

Efectividad, seguridad y coste-efectividad de los sistemas automatizados de detección de reacciones adversas a medicamentos

Effectiveness, safety and cost-effectiveness of automated systems for detecting adverse drug reactions

Informes de Evaluación de Tecnologías Sanitarias
SESCS

INFORMES, ESTUDIOS E INVESTIGACIÓN



MINISTERIO
DE SANIDAD



Efectividad, seguridad y coste-efectividad de los sistemas automatizados de detección de reacciones adversas a medicamentos

Effectiveness, safety and cost-effectiveness of automated systems for detecting adverse drug reactions

Informes de Evaluación de Tecnologías Sanitarias
SESCS

INFORMES, ESTUDIOS E INVESTIGACIÓN



MINISTERIO
DE SANIDAD



RED ESPAÑOLA DE AGENCIAS DE EVALUACIÓN
DE TECNOLOGÍAS Y PRESTACIONES DEL SISTEMA NACIONAL DE SALUD



Gobierno
de Canarias

RIVERO SANTANA A.

Efectividad, seguridad y coste-efectividad de los sistemas automatizados de detección de reacciones adversas a medicamentos/ A. Rivero Santana et al. – Madrid: Ministerio de Sanidad. Santa Cruz de Tenerife: Servicio Canario de la Salud. – 185 p.; 24 cm. – (Colección: Informes, estudios e investigación / Ministerio de Sanidad) (Serie: Informes de Evaluación de Tecnologías Sanitarias / SESCS)

NIPO: 133-24-103-8

1. Sistemas Automatizados 2. Detección 3. Eventos adversos 4. Medicamentos

I. Canarias. Servicio Canario de la Salud II. España. Ministerio de Sanidad

El Servicio de Evaluación de la Dirección del Servicio Canario de la Salud asume la responsabilidad exclusiva de la forma y el contenido final de este informe. Las manifestaciones y conclusiones de este informe son las del Servicio de Evaluación y no las de sus revisores externos.

Edita: Ministerio de Sanidad. Servicio Canario de la Salud

Este documento ha sido realizado por el Servicio de Evaluación del Servicio Canario de la Salud en el marco de la financiación del Ministerio de Sanidad para el desarrollo de las actividades del Plan anual de Trabajo de la Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del SNS, aprobado en el Pleno del Consejo Interterritorial del SNS el 23 de junio de 2023

Para citar este informe:

Rivero-Santana, A., Ramos-García, V., Delgado-Rodríguez, J., Torres-Castaño, A., Valcárcel-Nazco, C González-Pacheco, H., Abt-Sacks, A., Toledo-Chávarri, A., Guirado-Fuentes, C., Herrera-Ramos, E. Cazaña-Pérez, V., García-Bello, M., Santos Álvarez, A., Duarte-Díaz, A., Pérez-Álvarez, Y., Ramallo-Fariña, Y., Masiero-Aparicio, P., Boada del Campo, C., López-Rodríguez, JA., Domingo-Torrell, L., Lugonez-Sánchez, C., Perestelo-Pérez, L. Efectividad, seguridad y coste-efectividad de los sistemas automatizados de detección de reacciones adversas a medicamentos. Ministerio de Sanidad. Servicio de Evaluación del Servicio Canario de la Salud; 2023. Informes de Evaluación de Tecnologías Sanitarias.



MINISTERIO
DE SANIDAD



RED ESPAÑOLA DE AGENCIAS DE EVALUACIÓN
DE TECNOLOGÍAS Y PRESTACIONES DEL SISTEMA NACIONAL DE SALUD



Gobierno
de Canarias

Agradecimientos

Los autores de este informe quieren expresar su agradecimiento a Carlos González Rodríguez por su apoyo en las tareas de documentación y edición.

Índice

SIGLAS Y ACRÓNIMOS	13
RESUMEN EN LENGUAJE NO ESPECIALIZADO	15
PLAIN LANGUAGE SUMMARY	17
RESUMEN EJECUTIVO	19
EXECUTIVE SUMMARY	26
I. INTRODUCCIÓN	33
I.1. Problema de salud	33
I.2. Descripción y características técnicas de la tecnología	36
I.3. Marco Regulatorio en España y situación actual de la tecnología en el Sistema Nacional de Salud	38
I.4. Justificación de la evaluación	40
II. OBJETIVOS	41
II.1. Objetivo general	41
II.2. Objetivos específicos	41
II.3. Preguntas de investigación	41
III. METODOLOGÍA	43
III.1. Efectividad y seguridad	43
III.1.1. Criterios de selección de estudios	43
III.1.2. Fuentes de información y estrategia de búsqueda	45
III.1.3. Proceso de selección de estudios	46
III.1.4. Evaluación crítica del riesgo de sesgo	46
III.1.5. Extracción de datos	46
III.1.6. Método de análisis y síntesis de los datos	47
III.1.7. Valoración de la calidad de la evidencia de efectividad y	

seguridad	47
III.2. Revisión sistemática de evaluaciones económicas	47
III.2.1. Criterios de selección de estudios	47
III.2.2. Fuentes de información y estrategia de búsqueda	48
III.2.3. Proceso de selección de estudios	49
III.2.4. Evaluación de la calidad metodológica	49
III.2.5. Extracción y síntesis de datos	50
III.3. Análisis económico	50
III.3.1. Estudio de costes	50
III.4. Análisis de aspectos éticos, legales, organizativos, sociales, y ambientales	51
III.4.1. Alcance de la revisión	51
III.4.2. Criterios de selección de estudios	53
III.4.3. Fuentes de información y estrategia de búsqueda y proceso de selección de estudios	53
III.4.4. Proceso de extracción de datos, análisis y síntesis de los hallazgos	54
III.4.5. Aspectos ambientales	55
III.5. Necesidades de investigación futura	56
III.5.1. Análisis de las necesidades de investigación identificadas en las RS realizadas para el presente informe	56
III.5.2. Revisión de los estudios específicos sobre necesidades de investigación y medidas de resultado estándares	56
III.7. Estimación de la fecha de actualización del informe	58
III.8. Participación de grupos de interés	59
III.8.1. Participación de pacientes	59
III.8.2. Participación de los profesionales sanitarios	59
III.8.3. Participación de la industria	60
III.8.4. Revisión externa	60
IV. RESULTADOS	61
IV.1. Efectividad y seguridad	61
IV.1.1. Resultados de la búsqueda bibliográfica	61
IV.1.2. Características de los estudios incluidos	62

IV.1.3. Riesgo de sesgo de los estudios incluidos.	69
IV.1.4. Efectividad y seguridad	71
IV.1.4. Calidad de la evidencia de efectividad y seguridad	79
IV.2. Revisión sistemática de evaluaciones económicas	79
IV.2.1. Resultado de la búsqueda bibliográfica	79
IV.3 Análisis económico	79
IV.3.1. Estudio de costes	79
IV. 4. Aspectos éticos, legales, organizativos, sociales y ambientales	80
IV.4.1. Resultados de las consultas con actores relevantes y expertos/as informáticos/as	80
IV.4.2 Resultados de la Revisión Sistemática	83
IV.4.3 Resultados de la revisión manual de la literatura dirigida	91
IV.5. Aspectos ambientales	93
IV.6. Necesidades de investigación futura e identificación de medidas de resultado estándares	93
IV.6.1 Necesidades de investigación identificadas en la RS realizada para el presente informe	94
IV.6.2 Necesidades de investigación identificadas en estudios específicos de necesidades de investigación	95
IV.6.3 Medidas de resultados estándares	95
IV.7 Estudios en marcha	95
V. DISCUSIÓN	97
V.1 Efectividad y seguridad	97
V.2 Coste-efectividad	99
V.3 Aspectos éticos, legales, organizativos, sociales y ambientales	99
V.4 Estimación de la fecha de actualización del informe	102
VI. CONCLUSIONES	103
CONTRIBUCIÓN DE LOS AUTORES Y REVISORES EXTERNOS	105
REPRESENTANTES DE ASOCIACIONES DE PACIENTES Y SOCIEDADES CIENTÍFICAS	110

DECLARACIÓN DE INTERESES	111
REFERENCIAS	113
ANEXOS	123
Anexo 1. Estrategias de búsqueda	123
Anexo 2. Estimación fecha de actualización del informe	135
Anexo 3. Estudios excluidos a texto completo	135
Anexo 4. Algoritmo para la realización de evaluación económica en Evaluación de Tecnologías Sanitarias	131
Anexo 5. Algoritmo para la realización de los aspectos éticos, legales, organizativos, sociales y ambientales en Evaluación de Tecnologías Sanitarias	132
Anexo 6. Listado de preguntas a los actores de interés	133
Anexo 7. Características de los estudios incluidos (aspectos éticos, legales, organizativos, sociales y ambientales)	146
Anexo 8. Valoración de las limitaciones metodológicas de estudios incluidos (aspectos éticos, legales, organizativos, sociales y ambientales)	153
Anexo 9. Características de los estudios incluidos (necesidades de investigación y medidas de resultado estándares)	159
Anexo 10. Marco de los principios de la OMS para garantizar una IA responsable	159
Anexo 11. Preguntas sobre aspectos ambientales realizadas a la industria	135
Anexo 12. Revisiones del informe	162

Índice de tablas

Tabla 1. Bases de datos consultadas (efectividad, seguridad y coste efectividad)	46
Tabla 2. Criterios de selección (evaluaciones económicas)	48
Tabla 3. Preguntas específicas de investigación.....	53
Tabla 4. Bases de datos consultadas (análisis de aspectos éticos, legales, organizativos, sociales y ambientales)	54
Tabla 5. Bases de datos bibliográficas consultadas (necesidades de investigación e identificación de medias de resultado estándares)	57
Tabla 6. Resultados de la búsqueda bibliográfica.....	61
Tabla 7. Características de los estudios primarios incluidos que evaluaron el rendimiento diagnóstico de los modelos desarrollados.	65
Tabla 8. Calidad metodológica de la revisión incluida (escala AMSTAR-2)	69
Tabla 9. Riesgo de sesgo de los estudios observacionales incluidos. Lista de verificación del Joanna Briggs Institute.....	70
Tabla 10. Resultados de los estudios primarios que evaluaron el rendimiento diagnóstico de los modelos desarrollados.	77
Tabla 11. Resultados de las consultas realizadas a los/as grupos de interés	80
Tabla 12. Resultados de la búsqueda bibliográfica (aspectos éticos, legales, organizativos, sociales y ambientales)	83
Tabla 13. Principales características de los estudios incluidos. Necesidades de investigación y medidas de resultado estándares, en el uso de sistemas automatizados de detección de reacciones adversas a medicamentos	161

Índice de figuras

Figura 1. Proceso de selección de estudios de efectividad y seguridad para sistemas de monitorización	62
Figura 2. Proceso de selección de estudios de aspectos éticos, legales, organizativos, sociales y ambientales	84

Siglas y acrónimos

AC	<i>Accuracy</i>
AEMPS	Agencia Española de Medicamentos y Productos Sanitarios
C.A.	Comunidad Autónoma
COMET	<i>Core Outcome Measures in Effectiveness Trials</i>
CPAF	Comisión de Prestaciones, Aseguramiento y Financiación
EA	Eventos adversos
EAM	Eventos adversos a medicamentos
ECA	Ensayo controlado aleatorizado
ELOSA	Ético, legal, organizacional, social y ambiental
EMA	Agencia Europea del Medicamento, del inglés <i>European Medicine Agency</i>
EO	Estudio observacional
ES	Especificidad
ETS	Evaluación de tecnologías sanitarias
FEDRA	Farmacovigilancia Española Datos de Reacciones Adversas
HCE	Historia Clínica Electrónica
IA	Inteligencia Artificial
ICHOM	<i>International Consortium for Health Outcomes Measurement</i>
INE	Instituto Nacional de Estadística
JBÍ	<i>Joanna Briggs Institute</i>
JLA	<i>James Lind Alliance</i>
MA	Meta-análisis
ML	<i>Machine learning</i>
OMS	Organización Mundial de la Salud

PLN	Procesamiento de Lenguaje Natural
RAM	Reacciones adversas a medicamentos
RCEI	Ratio de coste-efectividad incremental
RedETS	Red Española de Agencias de Evaluación de Tecnologías Sanitarias
RR	Riesgo relativo
RS	Revisión sistemática
SEFV-H	Sistema Español de Farmacovigilancia de medicamentos de uso Humano
SN	Sensibilidad
SNS	Sistema Nacional de Salud
UE	Unión Europea
VPP	Valor predictivo positivo

Resumen en lenguaje no especializado

Las reacciones adversas a medicamentos (RAM) son aquellos efectos dañinos, no deseados ni intencionados, que pueden ser tanto esperados como inesperados, que se producen a causa de un medicamento. Las RAM pueden ocurrir por errores en la prescripción de la medicación, por sobredosis o por un uso del medicamento para el que no está indicado, pero también pueden aparecer cuando se toma un medicamento de forma adecuada y en la dosis correcta. Las RAM son una causa frecuente de enfermedad, discapacidad, hospitalización e incluso muerte en todo el mundo, por lo que es importante que desde los sistemas sanitarios y las administraciones públicas se planifiquen estrategias para vigilar la ocurrencia de RAM en las personas que toman medicamentos. Actualmente, la estrategia establecida en muchos países, incluida España, es la notificación espontánea por parte tanto de personal sanitario como de la ciudadanía. Cuando un profesional sanitario sospecha que un paciente puede estar padeciendo una RAM, puede rellenar un formulario describiendo el caso, y esta información es evaluada por expertos en farmacia y farmacología clínica, y en última instancia es reportada a las autoridades regulatorias nacionales en materia de medicamentos, quienes evalúan esta información, la almacenan en la base de datos FEDRA (Farmacovigilancia Española de Datos de Reacciones Adversas), y monitorizan de forma continua la seguridad de los medicamentos analizando si es preciso tomar medidas.

Sin embargo, se ha comprobado que esta estrategia de notificación espontánea solo permite detectar una pequeña parte de todas las RAM que se estima que ocurren. Por eso, se han planteado otras estrategias como por ejemplo la búsqueda activa y sistemática de información en la historia clínica electrónica (HCE) de los pacientes sobre síntomas y enfermedades que pudieron ser provocadas por medicamentos, dado que toda esa información se encuentra en la HCE. Sin embargo, hacer este trabajo de forma manual requiere mucho tiempo por parte de profesionales sanitarios, incluyendo los de farmacia, por lo que no es posible aplicarlo masivamente. Para solventar este problema, se están desarrollando sistemas informáticos que permitan buscar automática y rápidamente en las HCE posibles casos de RAM. Actualmente, muchas de estas propuestas se basan en el aprendizaje automático (del inglés *machine learning*), una rama de la inteligencia artificial (IA) que ofrece

grandes posibilidades no solo en la detección de RAM, sino en muchas otras áreas de la medicina. Estos sistemas tienen la capacidad de aprender diferentes tareas y analizar una gran cantidad de información rápidamente. Dado que una gran parte de la información en la HCE está escrita por profesionales sanitarios con sus propias palabras, en lenguaje natural, un gran reto para los sistemas de IA es poder procesar estas notas para poder analizarlas y detectar posibles RAM en ellas. Esta tarea, el procesamiento de lenguaje natural (PLN), resulta de gran dificultad para los sistemas informáticos, dada la complejidad del lenguaje humano.

Este informe tiene como objetivo revisar los estudios científicos que investigan si estos sistemas realmente son útiles a la hora de detectar las RAM y reducir sus efectos negativos sobre la salud de las personas. También se han evaluado aspectos sobre la percepción y expectativas que los profesionales tienen sobre estos sistemas y cómo afectaría a su trabajo, su facilidad o dificultad de implantación y el coste que supondría implantar uno de estos sistemas en un centro sanitario.

Los resultados muestran que no existen estudios actuales donde se haya evaluado la detección de RAM con datos en vida real mediante alguno de estos sistemas. Tan solo se identificó un estudio realizado en EE.UU. que implantó un sistema de PLN en un hospital durante 3 meses. Los resultados mostraron que los médicos recetaron menos medicación antihipertensiva que el sistema detectaba como sospechosa de estar causando RAM antes el ingreso. Sin embargo, no se evaluó si las RAM alertadas realmente lo eran o si se pudieron reducir sus efectos nocivos.

Aparte de este estudio, se identificaron otros 20 estudios (más otros 29 incluidos en una revisión realizada por otros autores) que evaluaron sistemas de detección de RAM, la mayoría de ellos mediante PLN. Sin embargo, estos estudios no evaluaron el efecto que tenía implantar un sistema de detección de RAM en un centro sanitario, sino una fase previa, consistente en comprobar su capacidad para procesar y clasificar correctamente las notas clínicas. Los resultados obtenidos son prometedores, pero aún existen problemas por resolver para alcanzar un rendimiento óptimo en la detección de posibles RAM en los textos que escriben los sanitarios en la HCE.

En cuanto al coste de implantar uno de estos sistemas en un centro sanitario, se estimó un coste inicial de 25 000 € de implantación y acompañamiento para realizar la integración del sistema, al que habría que añadir un coste anual de 45 000 € para mantenimiento, actualización de la base de datos y los costes propios de la infraestructura en la nube.

Plain language summary

Adverse drug reactions (ADRs) are those harmful, unwanted or unintended, both expected and unexpected effects that occur due to a medication. ADRs can occur due to errors in medication prescription, overdose or use of the medication for which it is not indicated, but they can also appear when a medication is taken appropriately and in the correct dose. ADRs are a frequent cause of illness, disability, hospitalization and even death throughout the world, so it is important that health systems and public administrations plan strategies to monitor the occurrence of ADRs in people taking medications. Currently, the established strategy in many countries, including Spain, is spontaneous notification by health professionals and citizens. When a healthcare professional suspects that a patient may be suffering from an ADR, they can fill out a form describing the case, and this information is reviewed by experts in pharmacy and pharmacology, along with the rest of the patient's clinical data, to determine if it is actually an ADR. This information is ultimately stored in the FEDRA database (Spanish Pharmacovigilance of Adverse Reaction Data).

However, it has been proven that this spontaneous reporting strategy only detects a small portion of all ADRs estimated to occur. For this reason, other strategies have been proposed, such as the active and systematic search for information in the electronic medical records (EHR) of patients about symptoms and diseases that could be caused by medications, given that all this information is found in the HCE. However, doing this work manually requires a lot of time on the part of experts, so it is not possible to apply it massively. To solve this problem, computer systems are being developed that make it possible to automatically and quickly search medical records for possible cases of ADR. Currently, many of these proposals are based on machine learning, a branch of artificial intelligence (AI) that offers great possibilities not only in the detection of AMR, but in many other areas of medicine. These systems have the ability to learn different tasks and analyze a large amount of information quickly. Given that a large part of the information in the EHR is written by healthcare professionals in their own words, in natural language, a great challenge for AI systems is to process these notes in order to analyze them and detect possible ADRs in them. This task, Natural Language Processing (NLP), is very difficult for computer systems, given the complexity of human language.

This report aims to review scientific studies that investigate whether these systems are really useful in detecting ADRs and reducing their negative effects on people's health. Aspects have also been evaluated regarding the perception and expectations that professionals have about these systems and how it would affect their work, their ease or difficulty of implementation and the cost that would be involved in implementing one of these systems in a healthcare center.

Unfortunately, the results show that there are no modern studies where the detection of real ADRs has been evaluated using any of these systems. Only one study was identified in the United States that implemented one of these systems in a hospital for 3 months. The results showed that doctors prescribed less antihypertensive medication than the system detected as suspected of causing ADR before admission. However, it was not evaluated whether the alerted ADRs were real or whether they were actually reduced.

Apart from this study, 20 other studies were identified (plus another 29 included in a review by other authors) that evaluated ADR detection systems, most of them using NLP. However, these studies do not evaluate the effect of implementing an ADR detection system in a health center, but rather a previous phase, consisting of checking its ability to correctly "read" clinical notes. The results obtained are promising, but there are still problems to be solved to achieve optimal performance in the detection of possible ADRs in the texts written by healthcare workers.

Regarding the cost of implementing one of these systems in a healthcare center, an initial cost of €25,000 was estimated for implementation and support to integrate the system, to which an annual cost of €45,000 for maintenance would have to be added updating the database and the costs of cloud infrastructure.

Resumen ejecutivo

Introducción

El daño a pacientes debido a eventos adversos (EA) de las intervenciones médicas es una causa relevante de muerte y discapacidad a nivel mundial. Dentro de los EA se incluyen los eventos adversos a medicamentos (EAM), que incluyen cualquier evento nocivo que ocurre a una persona mientras está en tratamiento farmacológico, independientemente de que se sospeche o no de dicho tratamiento como causa del evento. Por su parte, las reacciones adversas a medicamentos (RAM) se definen como toda aquella respuesta nociva, no deseada y no intencionada que se produce tras la administración de un fármaco, y existe una sospecha de causalidad. Esta definición no solo incluye los efectos nocivos e involuntarios derivados del uso autorizado de un medicamento en dosis normales, sino también aquellos relacionados con errores de prescripción y usos al margen de los términos de la autorización de comercialización, tales como la indicación fuera de ficha técnica, el uso equivocado, abuso del medicamento y la sobredosis. Las RAM son una causa frecuente de enfermedad, discapacidad o incluso muerte. Entre los factores que aumentan el riesgo de que se produzcan RAM se encuentran la polimedicación, la edad de los pacientes, las patologías previas como la insuficiencia hepática y/o renal, principales vías de metabolización y eliminación, y los factores genéticos.

Dada la comprobada infranotificación producida por los sistemas de notificación espontánea, la dificultad y el alto coste de revisar manualmente la historia clínica electrónica (HCE) de forma sistemática, se requieren sistemas automatizados de detección de RAM que faciliten la farmacovigilancia. Los sistemas de inteligencia artificial (IA) pueden ser de gran utilidad en esta tarea, especialmente los sistemas de procesamiento del lenguaje natural (PLN), pues gran parte de la información relevante para la detección de RAM en la HCE se encuentra en formato narrativo.

Objetivos

- Evaluar la efectividad y seguridad de los sistemas automatizados para la detección de RAM.

- Evaluar el coste-efectividad de los sistemas automatizados para la detección de RAM.
- Evaluar los costes de los sistemas automatizados para la detección de RAM desde la perspectiva del Sistema Nacional de Salud (SNS).
- Identificar las consideraciones éticas, legales, organizativas, sociales y ambientales relacionadas con la tecnología.
- Identificar las necesidades de investigación y medidas de resultado estándares desde la perspectiva de pacientes, familiares/personas cuidadoras, profesionales sanitarios/as y personal investigador sobre el uso de los sistemas de detección de RAM.

Método

Revisión sistemática de efectividad y seguridad

Se realizó una búsqueda de la literatura en las bases de datos MEDLINE, Embase, CDRS/CENTRAL y *Web of Science Core Collection* hasta junio de 2023. En la búsqueda no se restringió por fecha de inicio, pero en la selección solo se incluyeron estudios posteriores a 2013 dado el rápido desarrollo de esta tecnología. Se incluyeron revisiones sistemáticas (RS), ensayos controlados aleatorizados (ECA) y estudios observacionales (EO) publicados en inglés y español que evaluaran la efectividad/seguridad o el rendimiento diagnóstico de los sistemas automatizados para la detección de RAM a partir de datos incluidos en la HCE (incluidos resultados reportados por los pacientes mediante cuestionarios), en atención primaria u hospitalaria. Se incluyen sistemas de computación clásica con datos estructurados (p.ej., códigos ICD-10, datos de laboratorio, medicación, etc.) y sistemas de IA con datos estructurados y no estructurados (textos narrativos).

Se excluyeron estudios de detección de señales en farmacovigilancia, estudios sobre predicción de RAM futuros, estudios que evalúan sistemas de alerta exclusivamente en el momento de la prescripción del medicamento, estudios centrados exclusivamente en errores de medicación, y estudios de rendimiento diagnóstico incluidos en las revisiones incluidas.

Se evaluó el riesgo de sesgo de los estudios de efectividad mediante las herramientas RoB-2 (ECA) y el listado de verificación del Joanna Briggs Institute (JBI) para estudios observacionales (EO), y se realizó una síntesis narrativa de los resultados.

Revisión sistemática de coste-efectividad

Se llevó a cabo una RS de evaluaciones económicas (EE) para evaluar el coste-efectividad de los sistemas automatizados para la detección de RAM, siguiendo la misma estrategia de búsqueda que para la RS de efectividad y seguridad. Se realizó la valoración de la calidad metodológica de las EE por dos revisoras independientes, mediante el listado de criterios de Drummond et al. y/o la Ficha de Lectura Crítica FLC 3.0 de OSTEBA. Se previó extraer los datos relacionados con la identificación del artículo, el diseño y metodología, y con los resultados del estudio (perspectiva, horizonte temporal, costes, efectividad, RCEI, umbral coste-efectividad y, análisis de sensibilidad), con especial atención a la calidad de las fuentes de información empleadas para el desarrollo de la evaluación económica y la fuente de su financiación. Se previó realizar una síntesis narrativa con tabulación de las características y los resultados de los estudios incluidos.

Estudio de costes

Se llevó a cabo una evaluación económica parcial (estudio de costes) desde la perspectiva del Sistema Nacional de Salud (SNS) de un sistema automatizado para la detección de RAM en personas con cualquier patología o condición clínica en la que haya indicación de tratamiento farmacológico. En base a un sistema actualmente en fase de pilotaje, se tuvieron en cuenta el uso de recursos y costes de la implantación y puesta en marcha de dicho sistema para interceptar el proceso de prescripción y poder realizar el soporte a prescripción farmacéutica.

Aspectos éticos, legales, organizativos, sociales y ambientales.

Se realizó una definición del alcance siguiendo el algoritmo para la evaluación de los aspectos éticos, legales, organizativos, sociales y ambientales específicos relativos a la tecnología. Para ello se realizaron una serie de búsquedas manuales y se consultó con las partes interesadas (autores/as y colaboradores/as del informe), así como con la industria. A partir del resultado de este proceso exploratorio se definieron las preguntas específicas de investigación, y se llevó a cabo una RS de la

literatura científica en las bases de datos electrónicas MEDLINE, Embase y Web of Science Core Collection. Se incluyeron revisiones sistemáticas y revisiones de alcance; revisiones narrativas, estudios cualitativos, estudios observacionales y estudios de métodos mixtos que cubriesen aspectos éticos, legales organizativos y sociales. Se realizó una síntesis narrativa de los resultados, y se complementó con nuevas consultas con las partes interesadas para clarificar aquellos aspectos que se consideraron relevantes.

Resultados

Efectividad/seguridad

Solo se identificó un estudio de efectividad. Se trata de un EO prospectivo (3 meses) con control histórico (3 meses previos) que evaluó los cambios en la prescripción de medicación antihipertensiva por parte de los médicos, tras la implantación de un sistema con un componente de PNL. Se observó una tasa significativamente menor de prescripción de medicación sospechosa de estar causando RAM (MSRAM) en comparación con el grupo control, durante las primeras 24 horas (28% vs. 39% de la MSRAM detectada, $p < 0.001$) y durante todo el ingreso (47% vs. 58%, $p < 0.001$). De la MSRAM prescrita durante el ingreso, el grupo de intervención mostró una tasa menor en las primeras 24 horas (77% vs. 84%, $p = 0.003$).

No hubo diferencias significativas en la tasa de discontinuación de la MSRAM antes del alta (16% vs 14%, $p = 0.40$). En la muestra global, no hubo diferencias en la tasa de suspensión/discontinuación de la MSRAM prescrita durante todo el ingreso ($p = 0.28$), pero sí cuando se analizaron solo los profesionales del grupo de intervención que respondieron al cuestionario post-alerta, es decir, quienes definitivamente vieron las alertas (25% vs. 15%, $p = 0.003$). En cuanto a la MSRAM mencionada en el informe de alta, la diferencia en suspensión/discontinuación no fue significativa en la muestra global (20% vs. 23%, $p = 0.053$), pero de nuevo sí lo fue al incluir solo a quienes respondieron el cuestionario (38% vs. 23%, $p < 0.001$).

Se incluyeron 20 estudios de rendimiento diagnóstico (17 con PLN), más otros 29 (todos de PLN) incluidos en la revisión incluida. Solo 7 de ellos verificaron la ocurrencia real de los eventos adversos incluidos en las bases de datos usadas. Del total de estudios de PLN, solo 6 realizaron una validación externa de sus modelos (i.e., con datos no usados en el entrenamiento). Los resultados en general muestran un buen rendimiento

global de los modelos (detección de todos los términos relevantes, medicación, síntomas, enfermedades, eventos adversos), pero el rendimiento es considerablemente menor cuando se trata de detectar en los textos los términos específicos (o las relaciones) que definen las posibles RAM. El rendimiento global varía ampliamente entre diferentes eventos adversos específicos evaluados dentro de un mismo estudio. Solo un estudio usó datos de cuestionarios de síntomas de forma prospectiva, obteniendo un alto rendimiento, aunque se trató de una pequeña muestra de pacientes sin validación externa.

Coste-efectividad

No se identificó evidencia sobre el coste-efectividad de los sistemas automatizados para la detección de RAM. No se pudo llevar a cabo una evaluación económica completa por carecer de la información mínima necesaria.

Estudio de costes

El estudio de costes desde la perspectiva del SNS estimó un coste inicial de 25 000 € de implantación y acompañamiento para realizar la integración del sistema al que habría que añadir un coste recurrente anual de 45 000 € para mantenimiento, actualización de la base de datos y los costes propios de la infraestructura en la nube.

Aspectos éticos, legales, organizativos, sociales y ambientales.

Respecto a los resultados referidos a los aspectos éticos y legales, cabe destacar la importancia de la privacidad en los sistemas automatizados de detección de RAM. El uso de grandes cantidades de datos (en inglés, Big Data) ha desempeñado un papel significativo en la identificación de RAM y en el intercambio de información sanitaria. Se deben explorar nuevas fuentes de datos, desarrollar herramientas de minería de datos y tecnologías de protección de la privacidad. Respecto a la ciberseguridad, los servicios de telemedicina en general suelen requerir conexiones de red inalámbricas que recogen y transmiten información de pacientes a los sistemas de telemedicina, permitiendo a terceras partes acceder a esos datos, por lo que el sistema de servicios de telemedicina puede estar expuesto a amenazas de seguridad. La existencia de sesgos potenciales en los conjuntos de datos utilizados para entrenar los algoritmos de sistemas de IA puede generar problemas de equidad.

Respecto a los aspectos organizativos, la evidencia obtenida indica que algunos profesionales consideran que los sistemas tecnológicos de notificación de RAM integrados en el flujo de trabajo pueden ser facilitadores para la notificación de RAM. Se identificaron algunos factores que afectan a la viabilidad de implementación de los sistemas automatizados, como la falta de estandarización entre sistemas de las HCE, la especificidad en las opciones de ingreso de datos, la documentación incompleta e inexacta, y la fatiga por alerta. Es importante que los sistemas automatizados para la notificación de los RAM estén integrados con la gestión de medicamentos o los registros electrónicos hospitalarios, ya que esto puede ayudar a ahorrar tiempo y recursos. La educación y una cultura positiva de notificación de RAM pueden ser útiles para aumentar las tasas de notificación, por lo que es necesario formar a los equipos de enfermería, medicina, farmacia y otros miembros del equipo sanitario.

Conclusiones

- No se ha identificado ningún estudio que evalúe la efectividad y seguridad de los sistemas automatizados de detección de RAM desarrollados en los últimos 10 años en la reducción de RAM y sus consecuencias sobre variables de salud finales. Solo un estudio evaluó la implementación hospitalaria de un sistema de PLN y su efecto sobre la prescripción de medicación antihipertensiva durante 3 meses, mostrando una reducción de la prescripción de medicación evaluada por el sistema como sospechosa de estar produciendo RAM previamente al ingreso.
- Se han encontrado un buen número de estudios que evalúan el rendimiento diagnóstico de los métodos de IA aplicados a los datos de HCE, especialmente PLN. Estos estudios muestran en general un buen rendimiento global de los modelos, pero que es debido principalmente a la detección de medicación, mientras que los resultados aún están por debajo de lo aceptable en la detección específica de RAM en los textos clínicos. Por otra parte, el rendimiento global muestra una amplia variabilidad cuando se estudian RAM específicas dentro de un mismo estudio. La mayoría de estos estudios solo realizaron una validación interna de sus modelos y no verificaron la ocurrencia

de RAM reales más allá de corroborar su mención en los textos para entrenar los modelos.

- La implementación de un sistema automatizado de detección de RAM debería llevarse a cabo únicamente tras una meticulosa planificación, en conjunción con los/as profesionales implicados, que tenga en cuenta el área clínica donde se maximice su efectividad y eficiencia (por ejemplo, en poblaciones con alto riesgo o en el caso de RAM graves), las expectativas y preocupaciones de los profesionales y el efecto sobre su flujo de trabajo, así como las cuestiones técnicas referidas a la viabilidad de su implementación, su actualización en el futuro, así como su escalabilidad y generalización a otras áreas clínicas y otros hospitales del sistema de salud.
- No se identificó evidencia sobre el coste-efectividad de los sistemas automatizados para la detección de RAM.
- El coste estimado de un sistema automatizado para interceptar el proceso de prescripción y poder realizar el soporte a prescripción farmacéutica fue de 25 000 €, en concepto de implantación y acompañamiento para realizar su integración en el Sistema de Información Hospitalaria. Para el mantenimiento, actualización de la base de datos y los costes propios de la infraestructura en la nube, habría que añadir un coste recurrente anual de 45 000 €.
- La planificación adecuada de la implementación de un sistema de detección precoz de RAM es un aspecto crucial para su introducción en el sistema sanitario. Dentro del contexto del SNS, es altamente recomendable incorporar los sistemas automatizados de detección de RAM a través de la oficina de interoperabilidad, que garantiza la calidad y seguridad del sistema, además de facilitar la implementación del mismo sistema en diferentes centros sanitarios. Adicionalmente, el uso de metodologías como el codiseño a la hora del desarrollo de estos sistemas automatizados puede responder a las necesidades expresadas por todas las partes interesadas.

Executive summary

Introduction

Harm to patients due to adverse events (AEs) from medical interventions is a relevant cause of death and disability worldwide. AEs include adverse drug events (ADEs), which include any harmful event that occurs to a person while they are receiving pharmacological treatment, regardless of whether or not said treatment is suspected as the cause of the event. For their part, adverse drug reactions (ADRs) are defined as any harmful, unwanted and unintentional response that occurs after the administration of a drug, and there is a suspicion of causality. This definition does not include harmful and non-intended effects derived from the authorized use of a medication in normal doses, but also prescription errors and uses at the margin of the terms of the marketing authorization, such as the indication of the technical prospect, the misuse, abuse of the medication and overdose. The ADR is a frequent cause of disease, disability or death. Among the factors that increase the risk of ADR are the polypharmacy, patients' age, previous pathologies like hepatic and/or renal insufficiency, main routes of metabolism and elimination, and genetic factors. Given the proven underreporting produced by spontaneous reporting systems and the impossibility of systematically manually reviewing the electronic health record (EHR), automated ADR detection systems are required to facilitate pharmacovigilance. Artificial intelligence (AI) systems can be very useful in this task, especially Natural Language Processing (NLP) systems, since much of the information relevant to the detection of ADR in the EHR is in narrative format.

Objectives

- To evaluate the effectiveness and safety of automated systems for the detection of ADR.
- To evaluate the cost-effectiveness of automated systems for the detection of ADR from the perspective of the National Health System (SNS).
- To estimate the costs of the incorporation of automated systems in electronic medical records would entail to the common portfolio of SNS services in Spain.

- To identify ethical, legal, organizational social and environmental considerations related to technology.
- To identify research needs and standard outcome measures from the perspective of patients, family members/caregivers, healthcare professionals and research staff on the use of ADR detection systems.

Methods

Systematic review on effectiveness and safety

A literature search was carried out in the MEDLINE, Embase, CDRS/CENTRAL and Web of Science Core Collection databases until June 2023. The search was not restricted by start date, but studies published after 2013 were excluded in the selection, given the rapid development of this technology. We included systematic reviews (SR), randomized controlled trials (RCTs) and observational studies (OS) that evaluated the effectiveness/safety or diagnostic performance of automated systems for the detection of ADRs from data included in the EMR (including patient-reported outcomes), in primary or hospital care. Classical computing systems with structured data (e.g., ICD-10 codes, laboratory data, medication, etc.) and AI systems with structured and unstructured data (narrative texts) are included.

We excluded studies of signal detection in pharmacovigilance, studies on prediction of future ADRs, studies that evaluate alert systems exclusively at the time of drug prescription, studies focused exclusively on medication errors, and studies of diagnostic performance included in the identified reviews.

A narrative synthesis of the results was carried out.

Systematic review on cost-effectiveness

An SR of economic evaluations (EE) was carried out to evaluate the cost-effectiveness of automated systems for the detection of ADRs following the same search strategy as for the SR of effectiveness and safety. The assessment of the methodological quality of the EE was carried out, by two independent reviewers, using the list of criteria of Drummond et al. and/or the Critical Reading Sheet FLC 3.0 of OSTEBA. It was planned to extract data related to the identification of the article, the design and

methodology, and the results of the study (perspective, time horizon, costs, effectiveness, ICER, cost-effectiveness threshold and sensitivity analysis), with special attention to the quality of the information sources used to develop the economic evaluation and the source of funding. A narrative synthesis was planned with tabulation of the characteristics and results of the included studies.

Cost study

A partial economic evaluation (cost study) was carried out from the perspective of the National Health System (SNS) of an automated system for the detection of ADRs in people with any pathology or clinical condition in which there is an indication for pharmacological treatment. Based on a system currently in the pilot phase, the use of resources and costs of the implementation and start-up of such system were taken into account to intercept the prescription process and be able to support pharmaceutical prescriptions.

Ethical, legal, organizational, social and environmental aspects

A definition of the scope was made following the algorithm for the evaluation of ethical, legal, organizational, social and environmental aspects specific to the technology. To this end, a series of manual searches were carried out and stakeholders (authors and contributors to the report) as well as industry were consulted. Based on the results of this exploratory process, specific research questions were defined, and an SR of the scientific literature was carried out in the electronic databases MEDLINE, Embase and Web of Science Core Collection. Systematic reviews and scoping reviews; narrative reviews, qualitative studies, observational studies and mixed methods studies covering ethical, legal, organizational and social aspects were included. A narrative synthesis of the results was made and supplemented by further consultation with stakeholders to clarify those aspects that were considered relevant.

Results

Effectiveness/safety

Only one effectiveness study was identified. This was a prospective OS (3 months) with historical control (previous 3 months) that evaluated changes in the prescription of antihypertensive medication by doctors, after the implementation of a system with an NLP component. A significantly lower

rate of prescription of medication suspected of causing ADR (MSADR) was observed compared to the control group, during the first 24 hours (28% vs. 39% of the MSADR detected, $p < 0.001$) and throughout the hospital stay (47% vs. 58%, $p < 0.001$). Of the MSADR prescribed during admission, the intervention group showed a lower rate in the first 24 hours (77% vs. 84%, $p = 0.003$).

There were no significant differences in the rate of MSADR discontinuation before discharge (16% vs 14%, $p = 0.40$). In the overall sample, there were no differences in the rate of suspension/discontinuation of the prescribed MSADR throughout the hospitalization ($p = 0.28$), but there were differences when only the professionals in the intervention group who responded to the post-alert questionnaire were analyzed, that is, those who definitely saw the alerts (25% vs. 15%, $p = 0.003$). Regarding the MSADR mentioned in the discharge report, the difference in suspension/discontinuation was not significant in the overall sample (20% vs. 23%, $p = 0.053$), but again it was significant when including only those who answered the questionnaire (38% vs. 23%, $p < 0.001$).

We included 20 diagnostic performance studies (17 with PLN), plus another 29 PLN studies included in the identified review. Only 7 of them verified the actual occurrence of the adverse events included in the databases used. Of the total NLP studies, only 6 performed an external validation of their models (i.e., with data not used in training). The results generally show a good overall performance of the models (detection of all relevant terms, medication, symptoms, diseases, adverse events), but the performance is considerably lower when it comes to detecting the specific terms (or relationships) in the texts that define the possible ADRs. Overall performance varies widely between different specific adverse events evaluated within the same study. Only one study used data from symptom questionnaires prospectively, obtaining high performance, although it was a small sample of patients without external validation.

Cost-effectiveness

No evidence was identified on the cost-effectiveness of automated systems for ADR detection. A complete economic evaluation could not be carried out due to lack of the minimum necessary information.

Cost study

The cost study from the perspective of the SNS estimated an initial cost of €25,000 for implementation and support to carry out the integration of the system, to which an annual recurring cost of €45,000 would have to

be added for maintenance, updating the database and the costs of cloud infrastructure.

Ethical, legal, organizational, social and environmental considerations

Regarding ethical and legal aspects, it should be noted that privacy is of utmost importance in automated ADR detection systems. The use of large amounts of data (Big Data) has played a significant role in ADR identification and health information sharing. New data sources should be explored, data mining tools and privacy protection technologies should be developed. Regarding cybersecurity, telemedicine services in general often require wireless network connections that collect and transmit patient information to telemedicine systems, allowing third parties to access these data, so the telemedicine service system may be exposed to security threats. The existence of potential biases in the datasets used to train AI system algorithms may generate fairness issues.

Regarding organizational aspects, the evidence suggests that some practitioners consider that ADR reporting technology systems integrated into the workflow can be facilitators for ADR reporting. Some factors affecting the feasibility of automated systems implementation were identified, such as lack of standardization between EHR systems, specificity in data entry options, incomplete and inaccurate documentation, and alert fatigue. It is important that automated systems for ADR reporting be integrated with medication management or hospital electronic records, as it can help save time and resources. Education and a positive culture of ADR reporting can be invaluable in increasing ADR reporting rates, so nursing, medical, pharmacy and other members of the healthcare team need to be trained to take responsibility for ADR reporting.

Conclusions

- No study has been identified that evaluates the effectiveness and safety of automated ADR detection systems developed in the last 10 years in reducing ADR and its consequences on final health outcomes. Only one study evaluated the hospital implementation of a NLP system and its effect on the prescription of antihypertensive medication for 3 months, showing a reduction in the prescription of medication evaluated by the system as suspicious of producing ADRs prior to admission.

- A good number of studies have been carried out evaluating the diagnostic performance of AI methods applied to EHR data, especially NLP. These studies generally show good overall performance of the models, but that is mainly due to the detection of medication, while the results are still below acceptable performance in the specific detection of ADRs in clinical texts. On the other hand, overall performance shows wide variability when specific RAMs are studied within the same study. Most of these studies only carried out internal validation of their models and did not verify the occurrence of actual ADRs beyond corroborating their mention in the texts to train the models.
- The implementation of an automated ADR detection system should be carried out only after meticulous planning, in conjunction with the clinical professionals involved, taking into account the clinical area where its effectiveness and efficiency are maximized (for example, in populations with high risk or in the case of severe ADRs), the expectations and concerns of professionals and the effect on their workflow, as well as the technical issues regarding the feasibility of its implementation, its updating in the future, as well as its scalability and generalization to other clinical areas and other hospitals in the health system.
- No evidence was identified on the cost-effectiveness of automated systems for ADR detection.
- The estimated cost of an automated system to intercept the prescription process and be able to support pharmaceutical prescriptions was €25,000, for implementation and support to integrate it into the Hospital Information System. For maintenance, updating of the database and the costs of the cloud infrastructure, an annual recurring cost of €45,000 would have to be added.
- Proper planning of the implementation of an early ADR detection system is a crucial aspect for its introduction into the healthcare system. Within the context of the NHS, it is highly recommended to incorporate automated ADR detection systems through the interoperability office, which guarantees the quality and safety of the system, as well as facilitating the implementation of the same system in different healthcare centers. Additionally, the use of methodologies such as co-design when developing these automated systems can respond to the needs expressed by all stakeholders.

I. Introducción

I.1. Problema de salud

La incidencia de eventos adversos (EA) derivados de intervenciones sanitarias constituye una significativa causa de morbilidad y mortalidad a escala global. Anualmente, los hospitales de países de ingresos bajos y medios experimentan una elevada incidencia de EA relacionados con una atención sanitaria insegura, alcanzando un total de 134 millones de casos, los cuales contribuyen a 2.6 millones de muertes, mientras que en los países de altos ingresos hasta un 10% de los pacientes hospitalarios sufre EA [1,2]. Dentro de los EA se incluyen los eventos adversos a medicamentos (EAM), que incluyen cualquier evento nocivo que ocurre a una persona mientras está en tratamiento farmacológico, independientemente de que se sospeche o no de dicho tratamiento como causa del evento. Por su parte, las reacciones adversas a medicamentos (RAM) se definen como toda aquella respuesta nociva, no deseada y no intencionada que se produce tras la administración de un fármaco, y existe una sospecha de causalidad [3]. Esta definición no solo incluye los efectos nocivos e involuntarios derivados del uso autorizado de un medicamento en dosis normales, sino también aquellos relacionados con errores de medicación y usos al margen de los términos de la autorización de comercialización, incluidos la indicación fuera de ficha técnica, el uso equivocado, o abuso del medicamento y la sobredosis [4]. Las RAM son una causa frecuente de enfermedad, discapacidad o incluso muerte [5]. Entre los factores que aumentan el riesgo de que aparezcan se encuentra la polimedicación, la edad de los pacientes, las patologías previas como la insuficiencia hepática y factores genéticos [6]. Además, factores psicológicos, como el estrés, la ansiedad, la depresión y otros trastornos mentales, pueden afectar la respuesta del cuerpo a los medicamentos, mediante desequilibrios hormonales o inmunológicos, lo que puede aumentar la sensibilidad a los efectos secundarios de los medicamentos [7].

Entre los tipos de RAM se encuentra las RAM tipo A (aumentado, del inglés *augmented*) que son el resultado de un aumento en la acción farmacológica del medicamento cuando se administra a la dosis terapéutica habitual. Normalmente son reacciones dosis-dependientes (p.ej., la baja presión arterial con antihipertensivos y la hipoglucemia con insulina). Por otro lado, las RAM tipo B (extraño, del inglés *bizarre*), son reacciones que no se esperan de las conocidas acciones farmacológicas

de la medicación (p.ej., la anafilaxia con penicilina, o las erupciones cutáneas con algunos antibióticos). Especialmente las RAM tipo B son uno de los objetivos prioritarios de la farmacovigilancia.

En el contexto sanitario, las RAM se asocian a una carga importante, en términos de costes directos asociados a la morbilidad, mortalidad, ingresos hospitalarios y tratamiento, entre otros. En Europa, se estima que aproximadamente el 5% de todos los ingresos hospitalarios en urgencias se deben a una RAM, produciendo 197 000 muertes al año y con un coste asociado para el sistema sanitario de 145 000 millones de euros [8].

En España, al igual que en otros países de la Unión Europea (UE), las RAM producen un coste medio incremental de 52 360€ y 11 905€, según el tipo de EA asociado, suponiendo un coste total de entre 6.7% y el 15.4% de los correspondientes episodios [9]. No obstante, estudios previos dirigidos a identificar la incidencia de RAM, han puesto en relieve cómo un número significativo de ingresos hospitalarios debidos a una RAM (entre el 2.4 y 6.5% del total de ingresos) son potencialmente prevenibles [10–12]. Así mismo, revisiones sistemáticas (RS) de la literatura han puesto de relieve que alrededor de 20-30% de estas RAM son prevenibles, bien porque son debidas a errores de medicación, por el uso de medicamentos inapropiados a la edad u otras características del paciente, o por omisión de alertas. Además, estas tasas de RAM prevenibles parecen mantener tendencias inalteradas en el tiempo, a pesar de las diferentes estrategias de salud implementadas a nivel de seguridad de los medicamentos [13–15].

Una RS de estudios observacionales publicada en 2018 concluyó que los costes debidos a RAM en el ámbito hospitalario tuvieron un rango más amplio (de 2851€ hasta un máximo de 9015€) que al coste ambulatorio (de 174€ hasta 8515€) [16]. En España, un estudio epidemiológico retrospectivo y descriptivo en el que se usó como fuente de datos el Conjunto Mínimo Básico de Datos del Ministerio de Sanidad entre los años 2001 y 2006, informó que los medicamentos más comúnmente asociados con la hospitalización de las RAM fueron los antineoplásicos e inmunosupresores, los corticoesteroides, los anticoagulantes y los antibióticos, y se informó de un aumento de costes de hasta el 19.1% para las RAM evaluadas [17]. En el último informe anual de la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), los medicamentos para trastornos generales (35 590), sistema nervioso (21 854) y sistema musculoesquelético (15 800) fueron las áreas con más notificaciones [18].

La base de datos EudraVigilance de la Agencia Europea del Medicamento (EMA, del inglés *European Medicine Agency*), tiene como objetivo gestionar y analizar la información sobre sospechas de RAM de medicamentos autorizados o en fase de estudio clínico en el Espacio Económico Europeo [19]. En España, se utiliza la notificación espontánea de sospechas de RAM, que consiste en comunicar, por parte de la ciudadanía y profesionales de la salud, las sospechas de RAM que observan durante su vida cotidiana o práctica clínica habitual al órgano competente en materia de farmacovigilancia de la Comunidad Autónoma (C.A.) correspondiente [20]. Dicha actividad es de obligado cumplimiento para cualquier profesional de la salud (RD 1344/2007) [21]. El Sistema Español de Farmacovigilancia de medicamentos de uso Humano (SEFV-H) es la estructura que integra las actividades de farmacovigilancia, incluida las notificaciones de profesionales y ciudadanía [4,18]. El SEFV-H cuenta con Centros Autonómicos de Farmacovigilancia ubicados en cada C.A., además del centro coordinador localizado en la División de Farmacoepidemiología y Farmacovigilancia de la AEMPS. Cada C.A. se responsabiliza de notificar sospechas de RAM en su territorio, registrándose en la base de datos FEDRA (Farmacovigilancia Española Datos de Reacciones Adversas) [22]. Las sospechas de RAM notificadas en FEDRA son incluidas, posteriormente, en EudraVigilance, así como en la base de datos de la Organización Mundial de la Salud (OMS) VigiBase.

La notificación se realiza mediante la “Tarjeta Amarilla”, un formulario que puede ser cumplimentado de forma electrónica o en papel y que incluye diversos campos básicos para cumplimentar y realizar la notificación de sospechas de RAM. Según el RD 1344/2007, cualquier profesional de la salud o persona usuaria puede notificar mediante esta vía, aunque para la ciudadanía no resulta de obligado cumplimiento [4]. El principal beneficio de este sistema de notificación espontánea es la posibilidad de aproximarse a una relación causal entre un acontecimiento adverso y el uso de un medicamento, cuando dicha relación causal se desconoce o existe poca información sobre ella, utilizando el algoritmo de Karch-Lasagna [23]. Sin embargo, una de sus principales limitaciones es la infranotificación, debida en parte, a la falta de tiempo para registrarlas o desconocimiento sobre el procedimiento, así como la imposibilidad de calcular la incidencia real de RAM [24–26]. Los sistemas no automatizados de notificación de RAM han evidenciado niveles de uso insuficientes, revelando diverso tipo de dificultades como: la falta de sensibilización del personal sanitario, restricciones temporales, carencia de información adecuada (lo que puede llevar a que las RAM notificadas espontáneamente sean de muy mala calidad o que se produzcan sesgos),

escasez de incentivos, implicaciones legales y la actitud hacia la notificación. Esta última se caracteriza por una motivación limitada y la carencia de sistemas de notificación eficaces y de fácil uso [27,28].

Para mejorar la efectividad y eficiencia de la farmacovigilancia, se han desarrollado sistemas informáticos para identificar de forma automática eventos sospechosos de representar una RAM a partir de los datos de la historia clínica electrónica (HCE) [29], mediante algoritmos de computación clásica o más recientemente mediante sistemas de inteligencia artificial (IA) [30], incluido el procesamiento del lenguaje natural (PLN) de textos narrativos incluidos en la HCE [31]. De esta forma pueden procesarse grandes cantidades de datos de forma efectiva, aunque la evidencia sobre el rendimiento de estos sistemas aun es incierta.

I.2. Descripción y características técnicas de la tecnología

La progresiva digitalización de la HCE ha posibilitado la implementación de algoritmos informáticos para buscar información de forma rápida y accesible, o incluir alertas para los profesionales sanitarios acerca de aspectos relevantes sobre la salud de sus pacientes. Sin embargo, el desarrollo de la detección automática de RAM durante los años 90 y la primera década de este siglo obtuvo resultados poco satisfactorios [29], debido a diferentes factores relacionados con limitaciones propias de la tecnología, imposibilidad de manejar datos no estructurados o la calidad de los sistemas de información subyacentes, además de otros factores relacionados con la propia investigación sobre las RAM (ej., definiciones y nomenclaturas de RAM no estandarizadas, heterogeneidad de las reglas de decisión, criterios de referencia inconsistentes).

En la última década, sin embargo, la mejora de los sistemas de información, así como la irrupción de la IA, han abierto nuevas y prometedoras perspectivas sobre el manejo de información incluida en la HCE. La IA, mediante técnicas de aprendizaje automático (ML, del inglés *machine learning*), muestra una gran potencialidad para mejorar la atención sanitaria en múltiples frentes, mejorando la precisión del diagnóstico, la medicina personalizada, la atención a pacientes y la administración de los servicios médicos en general [32,33]. En lo que se refiere a la gestión de datos clínicos, la IA ayuda a analizar grandes volúmenes de datos, identificando patrones y tendencias que pueden

guiar decisiones médicas, lo que puede contribuir a una mejora de la eficiencia en la administración de registros y flujos de trabajo.

Los modelos de IA pueden aplicarse al análisis de datos estructurados presentes en la HCE, como códigos diagnósticos, medicación, o datos de laboratorio. Sin embargo, una gran parte de la información registrada en la HCE se realiza mediante textos narrativos no estructurados por parte del personal sanitario, que incluyen información de gran importancia sobre diagnóstico, tratamiento y respuesta clínica del paciente. Los modelos de PLN se han desarrollado para analizar este tipo de información. El PLN es una rama de la IA que se ocupa de la interacción entre las computadoras y el lenguaje humano. Su objetivo es permitir que las máquinas procesen, generen y manipulen el lenguaje natural de manera efectiva. En el contexto de la atención sanitaria, el PLN puede desempeñar un importante papel en el procesamiento y análisis de la información contenida en las historias clínicas, que en gran parte se registra como textos narrativos que muchas veces conllevan una mayor riqueza informativa que los datos estructurados. Entre las aplicaciones del PLN se encuentran: 1) Reconocimiento de información clave de narrativas médicas, como identificar diagnósticos, tratamientos, medicamentos y otros aspectos importantes en los registros clínicos, como posibles RAM; 2) Normalización de entidades médicas: las historias clínicas a menudo contienen términos médicos variados. El PLN ayuda a normalizar estos términos, mapeándolos a conceptos estandarizados, lo cual facilita la búsqueda y el análisis; 3) Codificación automática: El PLN puede codificar automáticamente la información de las historias clínicas en formatos estructurados, como códigos de diagnóstico o procedimientos. Esto mejora la interoperabilidad y la gestión de datos; y 4) Resumen de texto clínico: Mediante técnicas de PLN, se pueden resumir extensos documentos clínicos en información concisa y relevante. Esto ayuda a lo/as profesionales a acceder rápidamente a los aspectos más importantes.

El entrenamiento de los modelos de IA puede hacerse mediante aprendizaje supervisado o no supervisado. En el primero es necesario un conjunto de datos previamente etiquetado (anotado), donde cada ejemplo tiene una etiqueta que indica la información relevante (p.ej., diagnósticos, medicamentos, etc.). Esta anotación suele hacerse manualmente por los profesionales expertos, aunque más recientemente se han desarrollado procedimientos semiautomáticos para realizar esta tarea. El sistema de aprendizaje supervisado utiliza estos datos etiquetados para aprender a reconocer patrones y extraer características relevantes del texto, como palabras clave, entidades médicas y estructuras gramaticales. Este tipo

de aprendizaje es el más usado en los estudios de PLN sobre RAM, pues puede lograr alta precisión en la clasificación de efectos adversos conocidos, y requiere menos datos para entrenar que el aprendizaje no supervisado, aunque el proceso de anotación (etiquetado) es costoso en términos de tiempo. Por su parte, el aprendizaje no supervisado agrupa documentos similares sin etiquetas previas, ayudando a descubrir patrones subyacentes en los datos. Es útil para agrupar medicamentos según similitudes en sus efectos adversos, y detectar patrones emergentes o relaciones inesperadas. Ambos tipos de aprendizaje requieren de un preprocesamiento de los datos, que en el caso del aprendizaje supervisado incluye pasos como la *tokenización* (dividir el texto en unidades más pequeñas, llamadas tokens), lematización y *stemming* (reducir las palabras a su raíz o forma base) o eliminación de *Stop Words* (descartar palabras comunes sin significado relevante).

Los modelos de PLN tratan de solventar dos tareas secuenciales: reconocimiento de entidades nombradas en los textos, es decir, reconocer los diferentes términos que se refieren a medicamentos, síntomas, diagnósticos, etc., y la extracción de relaciones entre dichos conceptos, en el caso de la detección de RAM, para determinar si se trata realmente de una RAM o de otra relación (p.ej., un síntoma de la enfermedad).

I.3. Marco Regulatorio en España y situación actual de la tecnología en el Sistema Nacional de Salud

La vigilancia y gestión de los riesgos asociados a los medicamentos, se realiza de forma coordinada a través del SEFV-H, integrado por la AEMPS, los centros autonómicos de farmacovigilancia y el personal sanitario. Es por esta razón que la colaboración de los/as profesionales sanitarios/as con la farmacovigilancia es una exigencia legal:

- La Ley 29/2006 de garantías y uso racional de los medicamentos y productos sanitarios, destaca la importancia de la farmacovigilancia como un elemento esencial del medicamento y en su artículo 53 establece la obligatoriedad de comunicar las sospechas de reacciones adversas por parte de los profesionales sanitarios. En el punto 3 del artículo 54 reafirma la obligatoriedad de colaboración de todos los profesionales sanitarios en tareas de farmacovigilancia.

- El Real Decreto 577/2013, establece la obligatoriedad a todo el personal de notificar las sospechas de reacciones adversas de los medicamentos autorizados, incluidas las de aquellos que se hayan utilizado en condiciones diferentes a las autorizadas. Estas notificaciones se enviarán lo más rápidamente posible al Centro Autonómico de Farmacovigilancia correspondiente, mediante cualquiera de las vías que éste ponga a su disposición. También las relacionadas con los medicamentos sujetos a un seguimiento adicional. Estos últimos estarán identificados mediante un símbolo, común para todo el ámbito de la UE, (triángulo negro invertido) que figurará en la información sobre el Medicamento. La lista de medicamentos sujetos a seguimiento adicional, está publicada en el portal web europeo de medicamentos y en el portal web de la Agencia Española de Medicamentos y Productos Sanitarios.
- Real Decreto Legislativo 1/2015, de 24 de julio, por el que se aprueba el texto refundido de la Ley de garantías y uso racional de los medicamentos y productos sanitarios.
- En marzo de 2024, el Parlamento Europeo aprueba la Ley de Inteligencia Artificial, cuyo objetivo es proteger los derechos fundamentales, la democracia, el Estado de derecho y la sostenibilidad medioambiental frente a la IA que entraña un alto riesgo, impulsando al mismo tiempo la innovación.

En el contexto español, existen iniciativas innovadoras para prevenir y detectar la aparición de RAM, como las desarrolladas en la Comunidad de Madrid (creación de la Historia Farmacoterapéutica, estandarización de la información farmacológica, módulo único de prescripción) [34], Canarias (desarrollo de una receta electrónica única visible desde atención primaria y especializada, y que incorpora información al profesional sobre: medicamentos de seguimiento adicional, excipientes, riesgo de posibles interacciones o fármacos inadecuados en ancianos) [35], Galicia (proyecto de innova-Saúde, que tiene como objetivo desarrollar un motor de reglas de decisión que permita identificar información y resultados relevantes dentro de los sistemas de historia clínica electrónica y generar alertas inteligentes por diferentes canales [36]), Navarra (programa “Observa” [37]), o Cataluña (módulo de seguridad clínica en la receta electrónica [38]).

En el ámbito del PLN aplicado a la detección de RAM, el grupo de investigación IXA de la Universidad del País Vasco ha publicado diversos trabajos de investigación básica, incluyendo la elaboración de varios *corpora* en Español (bases de datos anotadas por expertos/as en farmacología y farmacovigilancia), a partir de HCE de los hospitales de Basurto y Galdakao-Usansolo, a partir de los cuales han desarrollado modelos de detección de RAM [39–41],. Sin embargo, aún no existen iniciativas de implementación de estos sistemas en la práctica habitual.

I.4. Justificación de la evaluación

Como se ha comentado, las RAM representan un problema de salud pública, con consecuencias potencialmente graves para la persona que los padece y altos costes para el sistema sanitario. Así mismo, un porcentaje relevante de estas RAM (20-30%) son potencialmente prevenibles, siendo esencial su detección para intensificar las estrategias de salud en términos de farmacovigilancia y de “cultura de seguridad”. La detección automática de RAM puede suponer un avance en términos de salud para la población y reducción de costes para el sistema sanitario. Desde la Comisión de Prestaciones, Aseguramiento y Financiación (CPAF) del Ministerio de Sanidad se solicita una revisión sobre la efectividad, seguridad y coste-efectividad de los sistemas automatizados para la detección temprana de RAM en personas con cualquier indicación de tratamiento farmacológico y, particularmente, en aquellas que precisen de varios fármacos.

Por estas razones se propone elaborar un informe de evaluación de tecnologías sanitarias, de cara a sustentar científicamente la posible incorporación de sistemas automatizados para la detección de RAM a la cartera de servicios comunes del SNS.

II. Objetivos

II.1. Objetivo general

Evaluar la evidencia científica sobre la efectividad, seguridad y coste-efectividad de los sistemas automatizados para la detección de RAM.

II.2. Objetivos específicos

- Evaluar la efectividad y seguridad de los sistemas automatizados para la detección de RAM.
- Revisar la evidencia publicada sobre el coste-efectividad y evaluar los costes de los sistemas automatizados para la detección de RAM desde la perspectiva del Sistema Nacional de Salud (SNS).
- Identificar las consideraciones éticas, legales, organizativas, sociales y ambientales relacionadas con la tecnología.
- Identificar las necesidades de investigación y medidas de resultado estándares desde la perspectiva de pacientes, familiares/personas cuidadoras, personal sanitario y personal investigador sobre el uso de los sistemas de detección de RAM.

II.3. Preguntas de investigación

Efectividad clínica/seguridad:

- ¿Cuál es la efectividad y seguridad de los sistemas automatizados de detección de RAM en población con cualquier patología o condición clínica en la que haya indicación de tratamiento farmacológico?

Coste-efectividad:

- ¿Es coste-efectiva la implantación de los sistemas automatizados de detección de RAM en población con cualquier patología o condición clínica en la que haya indicación de tratamiento farmacológico? ¿Cuáles son los costes de su implantación en España?

Aspectos éticos, legales, organizativos, sociales y ambientales:

- ¿La introducción de los sistemas automatizados de detección de RAM en población con cualquier condición clínica en la que haya indicación de tratamiento farmacológico daría lugar problemas éticos, legales, organizativos o sociales? ¿Cuál sería el impacto ambiental que esta tecnología podría tener?

Necesidades de investigación futura

- ¿Cuáles son las necesidades de investigación y medidas de resultado estándares relacionadas con los sistemas automatizados de detección de RAM?

III. Metodología

Para abordar los objetivos del presente informe se realizó una RS de la literatura disponible sobre la efectividad, seguridad, coste-efectividad y aspectos éticos, legales, organizativos, sociales y ambientales.

III.1. Efectividad y seguridad

Se realizó una RS de la literatura, de acuerdo con la metodología desarrollada por la Colaboración Cochrane [42] y presentada siguiendo las directrices de la declaración PRISMA [43].

III.1.1. Criterios de selección de estudios

Se seleccionaron estudios originales que cumplieron los siguientes criterios de inclusión y exclusión:

III.1.1.1. Diseño de estudio

- RS con meta-análisis (MA). Solo se incluirán RS sin MA en caso de que aporten información no incluida en las primeras.
- Ensayos controlados aleatorizados (ECA).
- Estudios observacionales (EO).

Se excluyeron revisiones no sistemáticas y estudios cualitativos.

III.1.1.2. Tipo de participantes

Personas con cualquier patología o condición clínica en la que haya indicación de tratamiento farmacológico.

III.1.1.3. Intervención

Sistemas automatizados para la detección de RAM a partir de datos incluidos en la HCE (incluidos resultados reportados por los pacientes mediante cuestionarios), en atención primaria u hospitalaria. Se incluyen sistemas de computación clásica con datos estructurados (p.ej. códigos

ICD-10, datos de laboratorio, medicación) y sistemas de IA con datos estructurados y no estructurados.

Se excluyen:

- Estudios de detección de señales en farmacovigilancia, cuyo objetivo es, a partir de bases de datos de notificación espontánea y literatura científica (y recientemente también de historias clínicas), identificar RAM asociados a cada medicamento (especialmente los inesperados o no descritos), pero no identificar pacientes individuales que pueden estar padeciendo RAM.
- Estudios que desarrollan modelos de predicción de la ocurrencia de RAM a partir de datos del paciente y la medicación.
- Estudios que evalúan sistemas de alerta exclusivamente en el momento de la prescripción del medicamento, a partir de los datos clínicos y otra medicación del paciente, alertando sobre la posibilidad de RAM, contraindicaciones o interacciones entre medicamentos.
- Estudios centrados exclusivamente en errores de medicación.
- Se excluirán los estudios de rendimiento diagnóstico incluidos en las revisiones incluidas.

III.1.1.4. Comparadores

- Sistemas de detección de RAM no automatizados (notificación espontánea, sistemas activos no automatizados).
- Otro sistema automatizado.

III.1.1.5. Medidas de resultado

Efectividad y seguridad

- Mortalidad por RAM.
- Hospitalizaciones por RAM y tiempo de hospitalización.
- Visitas a urgencias.

- Número de RAM detectados.

Medidas de rendimiento

- Sensibilidad (SN) y especificidad (ES).
- Valor predictivo positivo (VPP).
- F-score (también llamada F1-score o F-measure): media armónica de SN y VPP.
- Área bajo la curva ROC.
- *Accuracy* (AC): precisión global (verdaderos positivos más verdaderos negativos / todos los casos). Dado que en la evaluación de modelos de IA el término inglés *precision* se usa como sinónimo de VPP, para evitar confusiones se mantendrá el término en inglés.

III.1.1.6. Tipo de publicación

Artículos originales publicados en revistas con revisión por pares. Se excluyeron editoriales, cartas al editor, opiniones y resúmenes de congresos.

III.1.1.7. Idioma de la publicación

Se incluyeron los estudios publicados en inglés y en español.

III.1.1.8. Fecha de publicación

En el protocolo no se determinó una fecha inicial para la inclusión de estudios, pero dado el rápido desarrollo de las tecnologías evaluadas se decidió incluir estudios publicados después de 2013.

III.1.2. Fuentes de información y estrategia de búsqueda

En la Tabla 1 se describen las bases de datos electrónicas consultadas desde su inicio hasta junio de 2023 y en el Anexo 1 se detallan las estrategias de búsqueda bibliográficas ejecutadas en cada una de ellas sin restricciones por idioma.

Tabla 1. Bases de datos consultadas (efectividad, seguridad y coste efectividad)

Base de datos	Plataforma de acceso	Periodo buscado
MEDLINE	Ovid	1946-2023
Embase	Elsevier	1972-2023
CDRS/CENTRAL	Cochrane Library - Wiley	1993-2023
Web of Science Core Collection	Clarivate	1900-2023

III.1.3. Proceso de selección de estudios

El proceso de selección de los estudios se realizó de forma independiente por dos revisores/as de forma paralela e independiente. En caso de duda y/o desacuerdo entre ellos, se acudió a un/a tercer/a revisor/a que comprobó los criterios del protocolo e intentó llegar a un consenso con los otros dos revisores. Las discusiones y los acuerdos quedaron documentados. La selección de los estudios se realizó a partir de los resúmenes recuperados en las bases de datos según los criterios de selección establecidos y previa valoración de la relevancia para este informe. A continuación, se recuperaron las publicaciones completas (incluyendo todas aquellas en las que su elegibilidad no se podía determinar por el resumen) y se volvió a comprobar la concordancia con los criterios de la revisión.

III.1.4. Evaluación crítica del riesgo de sesgo

La evaluación del riesgo de sesgo fue realizada de forma independiente por dos revisores/as; las discrepancias fueron resueltas por consenso o con la ayuda de un/a tercer/a revisor/a. Las RS se evaluaron con la escala AMSTAR-2 [44], los ECA con la herramienta RoB-2 desarrollada por la Colaboración Cochrane [45], mientras que los EO de efectividad con el instrumento correspondiente del *Joanna Briggs Institute* (JBI) [46]. No se evaluó la calidad metodológica de los estudios de rendimiento diagnóstico de forma individual, aunque se discutieron sus limitaciones conjuntamente.

III.1.5. Extracción de datos

La extracción de datos de los estudios incluidos fue llevada a cabo por un/a revisor/a y comprobada por un/a segundo/a revisor/a. Cuando hubo

desacuerdo entre ambos se resolvió tras discusión y cuando no hubo consenso se consultó con un/a tercer/a revisor/a. Las discusiones y los acuerdos quedaron documentados. Se diseñó una hoja de extracción de datos (en formato Excel) que recogía datos relacionados con la identificación del artículo (autores, fecha de publicación, país donde se realizó el estudio, etc.), el diseño y metodología (tipo de estudio, duración del estudio, características de las intervenciones en comparación, características de la población), y con los resultados del estudio (medidas de efectividad y medidas de seguridad).

III.1.6. Método de análisis y síntesis de los datos

Los resultados se describieron narrativamente, clasificando los estudios según el área clínica estudiada. No se analizó el sesgo de publicación.

III.1.7. Valoración de la calidad de la evidencia de efectividad y seguridad

La calidad de la evidencia fue valorada de forma cualitativa en el apartado de discusión.

III.2. Revisión sistemática de evaluaciones económicas

Se realizó una RS de evaluaciones económicas (EE) para evaluar el coste-efectividad de los sistemas automatizados para la detección de RAM, siguiendo la metodología desarrollada por la colaboración Cochrane [45] y presentado la información relativa a ella de acuerdo a las directrices de la declaración PRISMA [43].

III.2.1. Criterios de selección de estudios

Se previó la selección de estudios que cumplieran los criterios de inclusión recogidos en la Tabla 2.

Tabla 2. Criterios de selección (evaluaciones económicas)

	Criterios de inclusión	Criterios de exclusión
Diseño de estudios	EE completas en paralelo a estudios primarios o consistentes en modelos económicos: análisis coste-beneficio, coste-utilidad, coste-efectividad, coste-consecuencia y minimización de costes. Asimismo, se consideró la posibilidad de incluir análisis coste-consecuencia y EE parciales realizadas en España, dependiendo de la cantidad y calidad de las EE identificadas.	EE parciales foráneas
Población	Personas con cualquier patología o condición clínica en la que haya indicación de tratamiento farmacológico	
Intervención	Sistemas automatizados para la detección de RAM (esperados e inesperados) a partir de datos incluidos en la historia clínica (incluyendo procesamiento del lenguaje natural y resultados reportados por el paciente).	
Comparador	Sistemas de detección de RAM no automatizados (notificación espontánea, sistemas activos no automatizados); Otro sistema automatizado	
Medidas de resultado	RCEI y/o coste y efectividad de todas las alternativas en comparación	
Tipo de publicación	Estudios completos publicados como artículos originales o como informes de evaluación de tecnologías sanitarias.	Resúmenes de congresos, editoriales, cartas al editor, artículos de opinión.
Idioma de la publicación	Inglés o español	Otro idioma
Fecha de publicación	Sin límite	

EE: Evaluaciones económicas; **RAM:** Reacciones adversas a los medicamentos; **RCEI:** Ratio coste-efectividad incremental

III.2.2. Fuentes de información y estrategia de búsqueda

Para la identificación de EE no se realizó una búsqueda específica, sino que se empleó la misma estrategia de búsqueda que para la RS de efectividad y seguridad que se puede consultar en el apartado III.1.2 del presente informe. En el caso de los estudios económicos, se utilizaron filtros de evaluaciones económicas validados para hacer más específica la búsqueda.

En el Anexo 1 del presente informe se puede consultar la estrategia de búsqueda empleada para cada una de las bases de datos. Posteriormente, se completó la búsqueda de estudios publicados con el examen manual de los listados de bibliografía de los estudios seleccionados y la comprobación en Google Scholar de los estudios que citan a los estudios seleccionados.

Las referencias bibliográficas recuperadas en las distintas bases de datos se importaron al gestor de Zotero 6.0.23 [47] donde se integraron para obtener un único archivo con todas las referencias que, posteriormente, se procesaron a través de Deduplick, una aplicación basada en IA para la eliminación de duplicados de forma transparente [48]. Por último, las referencias únicas se exportaron a un documento de Microsoft Excel para realizar la selección de los estudios pertinentes, donde finalmente se hizo una verificación manual para comprobar si existía algún duplicado no detectado previamente.

De forma complementaria, se consultó la base de datos *Cost-Effectiveness Analysis (CEA) Registry* (fecha de búsqueda: 21/06/2023).

III.2.3. Proceso de selección de estudios

Dos revisoras (economistas) seleccionaron las referencias a partir de la lectura de los títulos y resúmenes localizados a través de la búsqueda en bases de datos bibliográficas. Se previó que los artículos seleccionados como relevantes fueran analizados a texto completo de forma independiente por las dos revisoras, que los clasificarían como incluidos o excluidos de acuerdo con los criterios de selección especificados. También se previó que las dudas o discrepancias fueran resueltas mediante consenso o con la ayuda de un tercer revisor, quedando las discusiones y los acuerdos documentados.

III.2.4. Evaluación de la calidad metodológica

Se previó la valoración, por dos revisoras independientes, de la calidad metodológica de los estudios incluidos mediante el listado de criterios de Drummond et al [49] y/o la Ficha de Lectura Crítica FLC 3.0 de OSTEBA (www.lectura critica.com) [50].

III.2.5. Extracción y síntesis de datos

Se previó extraer los datos relacionados con la identificación del artículo (autores, fecha de publicación, país donde se realizó el estudio, financiación, etc.), con el diseño y metodología, y con los resultados del estudio (perspectiva, horizonte temporal, costes, efectividad, RCEI, umbral coste-efectividad y, análisis de sensibilidad), con especial atención a la calidad de las fuentes de información empleadas para el desarrollo de la evaluación económica y la fuente de su financiación. Se previó realizar una síntesis narrativa con tabulación de las características y los resultados de los estudios incluidos.

III.3. Análisis económico

Se realizó un estudio de costes de un sistema automatizado para la detección de RAM en personas con cualquier patología o condición clínica en la que haya indicación de tratamiento farmacológico, desde la perspectiva del sistema sanitario. En base al algoritmo de SESCS-RedETS para la elaboración de EE, que se encuentra en el Anexo 2, no se consideró pertinente llevar a cabo una EE completa debido a la falta de la información mínima necesaria.

III.3.1. Estudio de costes

Se estimaron los costes directos sanitarios relacionados con un sistema automatizado para la detección de RAM. Como primer paso, se identificaron los recursos necesarios para la implantación y puesta en marcha de este sistema en un hospital del SNS. En un segundo paso, se estimó el uso de estos recursos en unidades naturales, en base a la información aportada por la empresa Costaisa, que está desarrollando un sistema de este tipo.

El sistema automatizado considerado consistió en un sistema que actualmente está en fase de pilotaje, con marcado CE del tipo *all-in-house* (en el que el desarrollo, mantenimiento y gestión de las infraestructuras, operaciones y funciones informáticas se realizan de manera interna). Corresponde a un servicio SaaS (*Software as a Service*, o Software como Servicio en español) integrado en el Sistema de Información Hospitalaria, pero en el que las aplicaciones estarían en la nube y serían accesibles a

través de un navegador web. La finalidad sería interceptar el proceso de prescripción para poder realizar el soporte a la prescripción farmacéutica.

III.4. Análisis de aspectos éticos, legales, organizativos, sociales y ambientales

III.4.1. Alcance de la revisión

La evaluación de los aspectos éticos, legales, organizacionales, sociales y ambientales específicos relativos a la tecnología se incluye en las definiciones recientes de ETS [51]. El análisis de dichos dominios se realizó aplicando el algoritmo para la evaluación de los aspectos éticos, legales, organizativos, sociales y ambientales específicos relativos a la tecnología (Anexo 3). Este algoritmo parte de una definición del alcance que valora si existen incertidumbres relevantes para la toma de decisiones sobre estos aspectos y permite explorar y conceptualizar el contexto de la intervención. Se entiende por incertidumbre relevante cuando se identifican con claridad preguntas de investigación específicas para la tecnología evaluada que deberían ser respondidas para evaluar adecuadamente los aspectos éticos, legales, organizacionales y ambientales específicos relativos a las tecnologías sanitarias en el contexto del SNS. Un enfoque pertinente para la evaluación de las tecnologías sanitarias es entender el contexto como relacional, dinámico y que emerge con el tiempo en múltiples niveles diferentes del sistema más amplio, en lugar de tratarse como un entorno externo en el cual se introduce una intervención.

Para esta definición del alcance se ha realizado un proceso exploratorio hermenéutico, es decir implicando una iteración interpretativa. Este enfoque es congruente con el carácter no directivo y flexible de la revisión de alcance que permite ampliar la obtención de fuentes de información a medida que se avanza en el conocimiento de la intervención a evaluar y de las preguntas de investigación iniciales [52]. En primer lugar, se ha realizado una serie de búsquedas manuales adoptando categorías analíticas del marco evaluativo del Core Model 3.0. de EUnetHTA (EUnetHTA 2016) [53] del marco GRADE *Evidence to Decision* (EtD) [54] para la toma de decisiones sobre sistemas sanitarios y salud pública y del marco “*VALues In Doing Assessments of healthcare TEchnologies*” (VALIDATE) [55]. La búsqueda se llevó a cabo en Google, Google Scholar, utilizando diferentes combinaciones de las palabras

clave combinando diferentes palabras de la estrategia de búsqueda llevada a cabo para la parte de efectividad con los términos de búsqueda propios de los aspectos explorados: *ethics, confidentiality, feasibility, acceptability, barriers, facilitators, implementation, equity, gender, qualitative, interview*. Esta búsqueda manual de alcance se llevó a cabo entre el 4/05/2023 y el 10/05/2023. De igual modo, se realizó una búsqueda manual específica para evaluar los aspectos ambientales utilizando diferentes combinaciones de los términos de la estrategia de búsqueda general referida a efectividad, más los términos pertinentes para esta dimensión, tales como *climate change, sustainability and greenhouse gases emissions*. Esta búsqueda manual realizó entre el 05/05/2023 y el 08/05/2023 y se repitió el 31/10/2023.

En segundo lugar, para esta definición del alcance, se consultó con las partes interesadas (autores/as y colaboradores/as del informe) y con la industria para solicitar información relativa los aspectos éticos, legales, organizacionales sociales, de pacientes y ambientales. Se realizó una consulta a través de una reunión online con un grupo de trabajo formado por un representante de asociaciones de pacientes, representantes de sociedades científicas, profesionales sanitarios (médicos y farmacéuticos) y representantes de la industria. Los/as participantes figuran en el apartado *Representantes de asociaciones de pacientes y sociedades científicas de este presente informe*. La consulta fue guiada por el equipo investigador de este informe mediante un guion semiestructurado disponible en el Anexo 4, que fue compartido previamente con las personas participantes potenciales. Esta reunión se celebró el día 20 de abril de 2023. Tras esta reunión se envió un resumen de la misma a los/as participantes, y también se les envió también un listado con las preguntas relevantes que habían surgido como resumen de los temas planteados en la reunión, pidiéndoles que priorizarán estas preguntas en base a su criterio de importancia.

De este proceso exploratorio se definieron las siguientes preguntas específicas de investigación (Tabla 3).

Tabla 3. Preguntas específicas de investigación

Pregunta	Resultados
1. ¿Cuáles son los principales obstáculos o dificultades para la implementación de estos sistemas? ¿Se puede implantar un único sistema a nivel estatal que permita una mejor armonización de los datos?	Factibilidad, barreras y facilitadores para la implementación y aceptabilidad
2. ¿Cómo gestionar las notificaciones para evitar saturación de alarmas en la historia clínica, desde la perspectiva de todos/as los/as profesionales involucrados: servicio de farmacia, medicina de familia, especialidades clínicas, ¿etc.?	Gestión eficaz de las alarmas/notificaciones del sistema
3. ¿Se puede llevar a cabo una implementación progresiva por grupos de riesgo en función del tipo de paciente, tipo de fármacos o en situaciones especiales, como la transición del alta hospitalaria al domicilio, de modo que además se permita identificar casos no sospechados ni por profesionales ni por pacientes?	Priorización de uso para una implementación progresiva
4. ¿Cuáles son los principales problemas relacionados con la ciberseguridad y confidencialidad que presentan los sistemas de detección automatizada de RAM?	Confidencialidad y ciberseguridad
5. ¿Existen sesgos en los sistemas automáticos que puedan tener repercusiones desde la perspectiva de la equidad?	Detección de sesgos en los sistemas de detección de RAM
RAM: Reacción adversa a medicamentos	

A partir de ahí, se llevó a cabo primeramente una RS de la literatura científica existente para tratar de responder a las preguntas identificadas en la fase anterior.

III.4.2. Criterios de selección de estudios de la RS

Se incluyeron revisiones sistemáticas y revisiones de alcance; revisiones narrativas, estudios cualitativos, estudios observacionales y estudios de métodos mixtos que cubriesen aspectos éticos, legales organizacionales y sociales. Únicamente se incluyeron los documentos escritos en español e inglés.

III.4.3. Fuentes de información, estrategia de búsqueda y proceso de selección de estudios

Se consultaron las siguientes bases de datos electrónicas: Embase y *Web of Science Core Collection* (Tabla 4). En el Anexo 1 se muestra la

estrategia de búsqueda utilizada en cada una de las bases de datos. No se aplicaron restricciones por fecha ni por idioma de publicación.

Dos revisoras seleccionaron por pares, de forma paralela e independiente, los estudios a partir de la lectura de los títulos y resúmenes localizados. Los artículos seleccionados como relevantes fueron analizados a texto completo por las dos revisoras de forma independiente, que los clasificaron como incluidos o excluidos de acuerdo con los criterios de selección especificados. Las revisoras contrastaron sus opiniones y, cuando hubo dudas o discrepancias, éstas fueron resueltas tras discusión.

Tabla 4. Bases de datos consultadas (análisis de aspectos éticos, legales, organizativos, sociales y ambientales)		
Base de datos	Plataforma de acceso	Fecha de búsqueda
MEDLINE	Ovid	02/06/2023
Embase	Elsevier	05/06/2023
Web of Science Core Collection	Clarivate Analytics	05/06/2023

III.4.4. Proceso de extracción de datos, análisis y síntesis de los hallazgos

Dos revisoras realizaron la extracción de datos de los estudios incluidos de manera independiente. Los datos extraídos fueron los relativos a la identificación del artículo (autores, fecha de publicación, país donde se realizó el estudio, etc.), metodología (diseño) y resultados del estudio sobre aspectos éticos, legales, organizativos y sociales relevantes para este informe. Estos datos fueron recogidos en una hoja electrónica en formato Excel diseñada *ad hoc*. Se evaluaron las limitaciones metodológicas de los estudios con escalas específicas para cada tipo de estudio: el instrumento CASPe [56] para los estudios cualitativos, SANRA [57] para las revisiones narrativas, el instrumento de lectura crítica *Mixed Methods Appraisal Tool* [58] para los estudios de métodos mixtos, la escala de valoración de la evidencia para series de casos del IHE y la lista de comprobación para la evaluación crítica de estudios transversales lista de verificación de la JBI [46].

Se realizó una síntesis narrativa teniendo en cuenta criterios de relevancia y coherencia de los resultados. La relevancia evalúa hasta qué punto el conjunto de las evidencias y argumentos analizados que apoyan un resultado de la revisión es aplicable al contexto específico de aplicación, teniendo en cuenta la población, el fenómeno de interés y el

entorno organizativo y social. La coherencia analizó la medida en la que los hallazgos estaban basados en los estudios analizados y proporcionaba una explicación razonable a los patrones encontrados en estos datos.

Entrevistas con participantes relevantes.

Dado que algunas de las preguntas identificadas en la fase de definición de alcance no pudieron ser respondidas con la evidencia disponible, y para complementar los resultados obtenidos en la RS y contextualizar la literatura en nuestro ámbito, se llevaron a cabo reuniones de seguimiento con el fin de recopilar información para poder situar más adecuadamente los aspectos éticos, legales, organizacionales y sociales relacionados con la tecnología. Concretamente, se contactó con diferentes técnicos/as informáticos/as con conocimientos sobre la tecnología, obteniendo una única respuesta.

Revisión manual de la literatura dirigida

Finalmente, se decidió llevar a cabo una búsqueda manual de la literatura para complementar los hallazgos de la RS, y para contextualizar la pregunta 5, a la que no se pudo responder en la RS. Concretamente, los artículos relativos a esta pregunta, no se refieren exclusivamente a los sistemas de IA desarrollados para la detección de RAM sino a sesgos existentes en los sistemas de IA aplicados al ámbito sanitario, dado que no se han identificado artículos relativos a estas cuestiones en la RS. Sin embargo, es importante tener en consideración los sesgos ya descritos en la literatura, así como las estrategias para minimizarlos, a la hora de valorar el contexto de implementación de la tecnología.

La búsqueda manual se llevó a cabo en Google, Google Scholar y Pubmed utilizando diferentes combinaciones de las palabras clave combinando diferentes palabras de los aspectos explorados: *“bias, artificial intelligence, sex, gender, racial, disparities, and equity”*. Esta búsqueda manual de alcance se llevó a cabo entre el 1/09/2023 y el 15/09/2023.

III.4.5. Aspectos ambientales

Para evaluar los aspectos ambientales relacionados con la tecnología, además de la búsqueda manual mencionada, se envió a través de correo electrónico (02/11/2023) un breve cuestionario a representantes de la

industria (ver Anexo 5), solicitando información sobre el impacto ambiental del desarrollo de los softwares para la implementación de sistemas automatizados de detección de RAM basados en IA y el impacto ambiental (en términos de consumo eléctrico y huella de carbono) del mantenimiento de los sistemas automatizados de detección personalizados de RAM basados en IA. Al no recibir respuesta, se envió un recordatorio (14/11/2023).

III.5. Necesidades de investigación e identificación de medidas de resultado

Este apartado tiene como objetivo: 1) identificar las necesidades de investigación actuales, es decir, las áreas de incertidumbre que requieren de más investigación por no disponer de evidencia científica y; 2) identificar las medidas de resultado consideradas como estándares o necesarias para medir de manera adecuada el efecto de una intervención y facilitar la comparación entre estudios.

En concreto, para este informe, se han buscado las necesidades de investigación y las medidas de resultado relacionadas con el uso de sistemas automatizados de detección de RAM desde el punto de vista de las diferentes partes interesadas. Para llevar a cabo estos objetivos se realizaron las siguientes fases:

III.5.1. Análisis de las necesidades de investigación identificadas en las RS realizadas para el presente informe

A lo largo del proceso de evaluación de efectividad, seguridad y coste-efectividad y de los aspectos éticos, de pacientes, sociales, legales, organizacionales y ambientales, se identificaron las áreas de incertidumbre que requieren de más investigación.

III.5.2. Revisión de los estudios específicos sobre necesidades de investigación y medidas de resultado estándares

Se realizó una revisión de los estudios identificados en la base de datos *James Lind Alliance (JLA)* que incluye estudios específicos sobre

necesidades de investigación. Para identificar las medidas de resultados estándares se acudió a las bases de datos específicas: a) *Core Outcome Measures in Effectiveness Trials* (COMET) y b) *International Consortium for Health Outcomes Measurement (ICHOM) Standards Sets* (Tabla 5).

La consulta a las fuentes específicas de información JLA, COMET y ICHOM se realizó utilizando los términos de búsqueda: “automated systems”, “patient safety”, “artificial intelligence”, “artificial neural network”, “machine learning”, “deep learning”, “artificial”, “machine”, “intelligence”, “learning”, “adverse events drugs”, “drug reaction”, “drug surveillance”, “pharmacovigilance”, “side effects”, “medical error”.

Tabla 5. Bases de datos bibliográficas consultadas (necesidades de investigación e identificación de medias de resultado estándares)		
Base de datos	Plataforma de acceso	Fecha de búsqueda
JLA	http://www.jla.nihr.ac.uk/	18/05/2023
COMET	http://www.comet-initiative.org/	18/05/2023
ICHOM	https://www.ichom.org/	18/05/2023

III.5.2.1. Alcance de la revisión

La revisión se centró en identificar las necesidades de investigación actuales desde el punto de vista de las partes implicadas (pacientes/familiares, personal sanitario, personal investigador y de gestión sanitaria) e identificar medidas de resultados estándares, sobre el uso de sistemas automatizados y seguridad de pacientes, ya que no se encontraron estudios específicos sobre el uso de las tecnologías a evaluar, en este caso, sistemas automatizados de RAM.

III.5.2.2. Criterios de selección de estudios

Para la identificación de necesidades de investigación y medidas de resultados estándares, se incluyeron estudios, tanto primarios como secundarios, que cubriesen específicamente estos aspectos referentes a la población, intervención y comparación mencionada.

Se excluyeron resúmenes de congresos, estudios en animales o estudios in vitro. Se incluyeron documentos en español e inglés y se excluyeron otros idiomas.

III.5.2.3. Proceso de selección de estudios, extracción de datos y síntesis de los hallazgos

La selección de estudios y extracción de datos fue realizada por una revisora a partir de la lectura de los títulos y resúmenes en primer lugar y, posteriormente, del texto completo de los estudios localizados a través de la búsqueda de la literatura. Las dudas se consultaron con una segunda revisora.

Se realizó una síntesis narrativa de los principales hallazgos encontrados y una breve descripción de las características de los estudios (autor principal, año, objetivo, diseño y población).

III.6. Estimación de la fecha de actualización del informe

En el marco actual del funcionamiento de RedETS, la fecha de actualización de cada informe de ETS se establece ocasionalmente y de forma individualizada a partir de la valoración conjunta entre las agencias evaluadoras, que proponen una fecha razonable de forma procedimentalmente no establecida aún; y la Secretaría de RedETS, que señala el grado de necesidad/prioridad política. Este informe de RedETS propone una fecha prevista de actualización, a partir de la consideración de criterios específicos; lo que permite una aproximación transparente y reproducible. Este ejercicio es especialmente relevante cuando, haciendo uso de la terminología GRADE, las conclusiones no permiten elaborar una recomendación, por requerirse de investigación adicional (recomendación para el uso únicamente en investigación); o la recomendación propuesta no es fuerte, independientemente de que su sentido sea favorable o desfavorable.

Los criterios concretos que, en este informe, fueron utilizados por el equipo evaluador para proponer una fecha de actualización pertinente, incluyen el ritmo de publicaciones observado durante los últimos años del período revisado y las fechas de publicación anunciadas por los posibles estudios en curso. Otros criterios complementarios, más difícilmente predecibles, y que quedan fuera del ámbito de actuaciones del evaluador, incluyen: 1) los cambios en los precios de las tecnologías o en los costes operativos, a propósito de cambios organizativos mediados por las TICs o por la reorganización funcional, que obligarían a actualizar tanto el análisis coste-efectividad, como los impactos presupuestario y organizativo; 2) cambios en la situación epidemiológica, expresados por

las variaciones en la prevalencia o mortalidad, que impactarían en los costes tanto por incrementar el número de personas usuarias como la utilización más intensa de los recursos sanitarios; además de 3) los posibles cambios en las indicaciones clínicas (Anexo 6).

III.7. Participación de grupos de interés

III.7.1. Participación de pacientes

La implicación de personas afectadas por las intervenciones a evaluar se planteó desde el inicio del proceso de evaluación con el objetivo de que pudieran realizar aportaciones sobre los aspectos relevantes para ellas. Se invitó a participar a las siguientes asociaciones y confederaciones de pacientes y familiares: Asociación Para las Prácticas Innovadoras en Polimedicación y Salud integrada en la Fundación para la Innovación, Investigación, Formación y el Desarrollo Comunitario (FÜNDEC), Asociación Española de afectados de cáncer de pulmón (AECAP), Federación Nacional de Asociaciones para la Lucha de las enfermedades del riñón (ALCER), Federación Nacional de Enfermos Trasplantados Hepáticos (FNETH), Federación Española de Diabetes (FEDE), Asociación de enfermos neuromusculares (Asem), Asociación Española Contra el Cáncer (AEC), Asociación Corazón Sin Fronteras, Federación Española de Pacientes Alérgicos y Con Enfermedades Respiratorias (FENAER), Federación de Asociaciones de Enfermería Familiar y Comunitaria (FAECAP), Confederación de Salud Mental. Se recibieron respuestas favorables de: FÜNDEC, AECAP, ALCER, FNETH y FEDE. Finalmente, cuatro revisaron el protocolo y/o la versión final del informe.

III.7.2. Participación de los profesionales sanitarios/as

Complementariamente, se difundió el propósito de realizar este informe entre las sociedades científicas cuya actividad profesional tuviera relación con el problema de salud. El equipo responsable de la elaboración de este informe invitó a participar a las siguientes sociedades científicas profesionales: Sociedad Española de Farmacología Clínica (SEFC), Sociedad Española de Farmacéuticos de Atención Primaria (SEFAP), Sociedad Española de Medicina de Familia y Comunitaria (semFYC), Sociedad Española de Medicina Interna (SEMI), Sociedad Española de Psiquiatría y Salud Mental (SEPSM), Asociación Española de Enfermería

Cardiológica (AEEC), Sociedad Española de Enfermería Oncológica (SEEO), Sociedad Española de Farmacia Hospitalaria, Sociedad Española de Medicamento Individualizado (LASEMI), Asociación Española de Farmacéuticos de la Industria (AEFI), Asociación Española de Enfermería de Salud Mental (AEESME), Asociación Española de Pediatría de Atención Primaria (AEPap), Sociedad Española de Enfermería Geriátrica y Gerontología (SEEGG). Se recibieron respuestas favorables por parte de las siguientes sociedades científicas: SEFC, SEFAP, FEFE, semFYC, SEMI, Sociedad SEPSM, AEEC, SEEO. Finalmente, seis revisaron el protocolo y/o la versión final del informe.

III.7.3. Participación de la industria

La comunicación de la realización de este informe también se difundió a la industria, a través de las asociaciones empresariales que integran a las empresas de tecnologías sanitarias (FENIN). Se invitó a participar a la empresa COSTAISA, a la Federación empresarial de farmacéuticos españoles (FEFE) y a la Asociación Española de medicamentos genéricos (AESEG).

III.7.4. Revisión externa

El presente informe, una vez finalizado y antes de su publicación, se sometió a un proceso de revisión crítica por parte de reconocidos/as expertos/as en el tema para asegurar su calidad, precisión y validez. A las personas revisoras se les solicitó que aportaran todos aquellos cambios o modificaciones que consideraran oportunos, debiendo estar sustentados en evidencia científica y acompañados de la bibliografía correspondiente. Los comentarios recibidos fueron valorados por el equipo autor del informe y estas valoraciones fueron remitidas a los/as revisores/as con las debidas justificaciones.

IV. Resultados

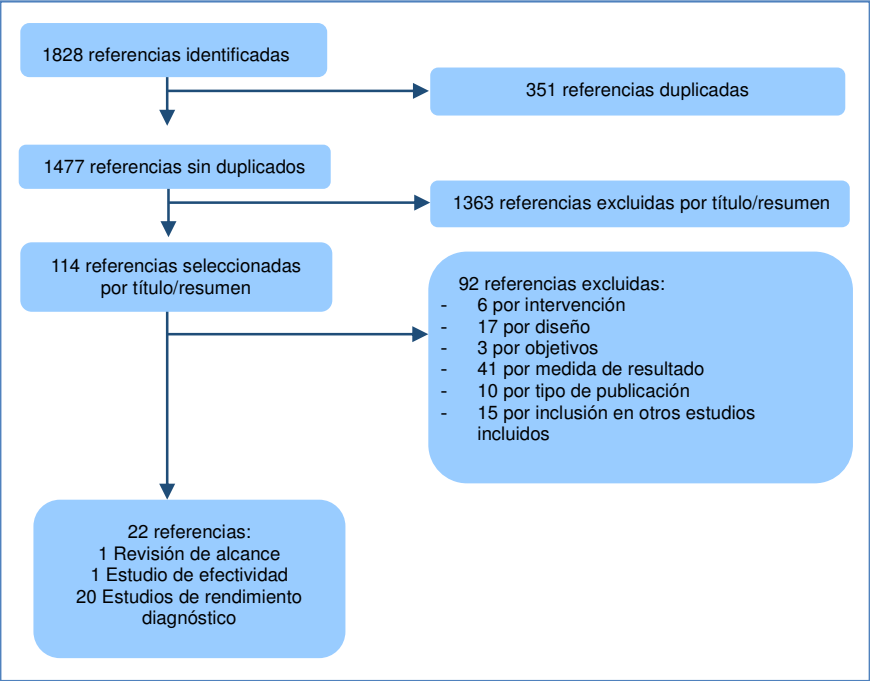
IV.1. Efectividad y seguridad

IV.1.1. Resultados de la búsqueda bibliográfica

El proceso de selección de estudios se muestra en la Figura 1. La búsqueda electrónica produjo 1828 referencias, 1477 tras eliminar duplicados (Tabla 6). Se seleccionaron 114 artículos por título y resumen, y tras la lectura de los artículos se incluyeron finalmente 22 estudios [31,41,59–78], una revisión de alcance [31], un estudio de efectividad [73] y 20 estudios de rendimiento diagnóstico. En el Anexo 7 se presentan los estudios excluidos.

Tabla 6. Resultados de la búsqueda bibliográfica				
Base de datos	Plataforma de acceso	Periodo buscado	Fecha de acceso	Nº de resultados
MEDLINE	Ovid	1946-2023	15/05/2023	645
Embase	Elsevier	1972-2023	15/05/2023	682
CDRS/CENTRAL	Cochrane Libray - Wiley	1993-2023	16/05/2023	31
Web of Science Core Collection	Clarivate Analytics	1900-2023	16/05/2023	470
TOTAL				1828
DUPLICADOS detectados por Software tras fusionar las búsquedas				347
DUPLICADOS detectados manualmente tras fusionar las búsquedas				4
TOTAL SIN DUPLICADOS				1477

Figura 1. Proceso de selección de estudios de efectividad y seguridad para sistemas de monitorización



IV.1.2. Características de los estudios incluidos

La revisión de alcance de Murphy et al. (2023) [31] incluyó estudios que evaluaran sistemas supervisados de PLN para detectar EAM en pacientes hospitalizados. Se excluyeron estudios en atención primaria o comunitaria, o que usaran fuentes de datos de farmacovigilancia, notificación espontánea o literatura científica. Se usaron las bases de datos MEDLINE, Embase y arXiv (base de datos de acceso abierto, con versiones iniciales de artículos aun no revisados por pares) hasta julio de 2021. Los resultados se describieron narrativamente, a partir de 6 dimensiones usadas como criterios de calidad metodológica (si bien no se evaluó la calidad de los estudios individuales): implicación de profesionales clínicos, descripción y preparación de los datos, modelado, evaluación del rendimiento e implementación en la práctica real.

Entre los estudios primarios, el único estudio de efectividad incluido es un estudio prospectivo con control histórico [73], realizado en un

hospital de Nashville (EEUU). Se evaluó un sistema de PLN (*Adverse Drug Effect Recognizer*, ADER) para la detección de RAM debidos a medicación previa al ingreso. Este sistema extrae información de los informes de ingreso sobre medicación, hallazgos clínicos y diagnosis previas, así como valores numéricos sobre resultados de laboratorio, y compara esta información con la incluida en una base datos sobre RAM conocidos desarrollada previamente por los autores. Pocos minutos después de introducirse una nota de admisión, el sistema identifica posibles RAM y genera una alerta (pasiva, para no interrumpir el flujo de trabajo), que incluye el posible EAM, la posible medicación causante, así como enfermedades del paciente que podrían estar causando el efecto (y por tanto no ser una RAM). Para los objetivos del estudio, se incluyó un cuestionario tras cada alerta para recoger información sobre ella por parte del médico.

El estudio se focalizó en RAM inducidos por medicación antihipertensiva. El sistema se implementó durante tres meses, y se recogió como variable de resultado, durante todo el ingreso y al alta, la tasa de prescripción de la medicación detectada por el sistema como sospechosa de estar induciendo RAM (MSRAM). Los 137 médicos y residentes participantes se compararon con otros 138 en las mismas unidades durante los tres meses previos al inicio del estudio. El sistema ADER revisó las notas de admisión del grupo control para determinar las alertas que habría recibido en ese periodo en caso de disponer del sistema, y definir la MSRAM en dicho grupo.

El resto de los estudios fueron EO de una cohorte en los que se evaluó el rendimiento diagnóstico de los modelos desarrollados. Sus características pueden verse en la Tabla 7. Todos se realizaron en el contexto hospitalario y usaron datos de sus propios centros o red de centros. Tres estudios se realizaron en Australia [64,68,69], EEUU [60,74,77] y Reino Unido [65,66,70], respectivamente, dos en España [41,72], Singapur [66,75] y Suecia [59,78] y uno en Corea del Sur [67], China [62], Finlandia [63], Francia [71] y Países bajos [76]. Todos usaron bases de datos retrospectivas, excepto el estudio de Ilvalanien et al. (2021) [63], que fue el único que usó datos de síntomas reportados por los pacientes mediante un cuestionario digital, aplicando un diseño prospectivo. Trece estudios evaluaron modelos de PLN mediante ML aplicado a diferentes tipos de información narrativa de la HCE (informes de admisión y alta, notas de enfermería, etc.) [41,60,64–68,70,72,74–77] y siete analizaron datos estructurados (datos de laboratorio, códigos ICD-10, medicación), cuatro de ellos mediante ML [59,63,69,78] y tres mediante algoritmos clásicos con reglas basadas en criterios clínicos

[61,62,71]. Dos de estos tres estudios [62,71] no usan una muestra de entrenamiento de la cual derivar o depurar el modelo, por lo que su validación se ha considerado externa. Siete estudios aplican un criterio de referencia *gold standard*, verificando la ocurrencia real de la RAM [60–63,69,71,78], aunque solo Chen et al. (2020) [62] describe los criterios de causalidad usados (*World Health Organization*, WHO). Entre los 17 estudios que aplicaron técnicas de IA, solo 5 aplicaron una evaluación externa [41,68,72,75,77], incluyendo datos de un hospital distinto en 4 de ellos [41,72,75,77]. Cinco estudios ofrecen resultados para más de un EAM/RAM específicos [59,65,66,76,78], mientras que el resto evaluó la detección de cualquier EAM/RAM (para cualquier medicación o para un número específico de ellas), dentro del ámbito de la población clínica estudiada.

Tabla 7. Características de los estudios primarios incluidos que evaluaron el rendimiento diagnóstico de los modelos desarrollados.

	Diseño	Población	EAM/RAM	Fuente de datos	Modelo	Criterio de referencia	Validación
General							
Bagattini et al. (2019) Suecia	ORnC	Pacientes de un hospital universitario (2009-2015)	15 EAM	Datos de laboratorio (1877 medidas)	Marco de clasificación de datos temporales (evaluado con RF)	Código ICD-10	Interna Cruzada (k10)
McMaster et al. (2019) Australia	ORnC	Ingresados en hospital terciario (2016-2017)	Cualquier RAM	Código ICD de RAM (2917 códigos) y de diagnóstico primaria; días de estancia	ML (varios modelos)	GS1: Farmacéutico/a experto/a y eventualmente reunión del grupo experto/a	Interna Partición (80%/20%)
McMaster et al. (2023) Australia	ORnC	Ingresados en hospital terciario (2015-2020)	Cualquier RAM	Informes de alta de admisiones con código ICD-10 de RAM (861)	PLN (pre-trained LLM)	Anotación por parte de un/a clínico/a, revisada por un/a clínico/a senior	Interna Cruzada (k5) Externa
Jeon et al. (2020) Corea del Sur	ORnC	Ingresados en hospital terciario (1994-2015)	Cualquier RAM	Notas de enfermería anteriores al RAM (8316)	PLN (varios modelos)	GS1: Informes de RAM	Interna Partición (80%/10%/10%)
Santiso et al. (2019) España	ORnC	Pacientes de un hospital (2008-2012)	Cualquier RAM	HCE (75 registros)	PLN (RF)	Anotación por 2 expertos/as en farmacia y farmacovigilancia	Interna Partición (70%/30%)
Santiso et al. (2021) España	ORnC	Pacientes de dos hospitales	Cualquier RAM	HCE (Interna: 75 registros; Externa: 463 registros)	PLN (Joint AB-LSTM y WE)	Anotación por 2 expertos/as en farmacia y farmacovigilancia	Interna Externa
Tang et al. (2019) Singapur	ORnC	Pacientes ingresados en un hospital (2011)	Cualquier EAM	Informes de alta (1620)	PLN (FL)	Tres anotadores/as independientes	Interna Partición 1020/600 Externa

Tabla 7. Características de los estudios primarios incluidos que evaluaron el rendimiento diagnóstico de los modelos desarrollados.

	Diseño	Población	EAM/RAM	Fuente de datos	Modelo	Criterio de referencia	Validación
Zhao et al. (2015) Suecia	ORnC	Ingresados en hospital universitario (2009-2010)	27 EAM	Datos de laboratorio y códigos ICD-10	ML (RF)	GS ¹ : Código ICD-10 de EAM con notificación explícita de causalidad	Interna Cruzada (k10)
Cáncer							
Iivanainen et al. (2021) Finlandia	OPnC	Pacientes con cáncer recibiendo IPCI	EAM por IPCI	HCE: datos clínicos estructurados y síntomas reportados por los pacientes; (34 pacientes, 840 cuestionarios)	ML (GBA)	GS ¹ : Confirmación del personal médico	Interna Partición (70%/30%)
van Laar et al. (2020) Países Bajos	ORnC	Pacientes con carcinoma renal metastásico tratados en hospital universitario (2015-2019)	Diarrea Hepatotoxicidad Hipertensión Síndrome mano-pie	HCE (99 pacientes)	PLN	Anotación por farmacéutico/a con experiencia clínica y en revisión de HCE	Interna
Quillet et al. (2018) Francia	ORnC	Pacientes con cáncer y TOD en un hospital (2014-2015)	RAM por TOD	Códigos ICD-10 tres meses después de iniciar la terapia	6 filtros aplicados a los códigos	GS ¹ : Revisión manual por centro de farmacovigilancia	Externa
Reacciones alérgicas							
Banerji et al. (2020) EEUU	ORnC	Ingresados en hospital universitario (2007-2008)	Reacciones alérgicas	HCE (3097 casos)	PLN en combinación con códigos ICD-10	GS ¹ : Investigador/a asistente entrenado/a y alergólogo/a senior	Interna Partición 81%/19%
Inglis et al. (2021) Australia	ORnC	Pacientes que toman penicilina en 3 hospitales (2013-2015)	Alergia vs. intolerancia a la penicilina	HCE (3963 casos)	PLN (varios modelos)	Anotación por un/a inmunólogo/a y un farmacólogo/a clínico/a	Interna Partición 75%/25%

Tabla 7. Características de los estudios primarios incluidos que evaluaron el rendimiento diagnóstico de los modelos desarrollados.

	Diseño	Población	EAM/RAM	Fuente de datos	Modelo	Criterio de referencia	Validación
Yang et al. (2020) EEUU	ORnC	Ingresados en 2 hospitales universitarios (2004-2019)	Reacciones alérgicas	Informes de seguridad (172 854 pacientes)	PLN (DNN)	Anotación por un/a investigador/a y un/a alergólogo/a o inmunólogo/a verificó la clasificación	BD1: cruzada (k5) BD3-4: externa
Psiquiatría							
Iqbal et al. (2017) Reino Unido	ORnC	Cuatro Hospitales psiquiátricos	19 EAM asociados a antipsicóticos y antidepresivos	HCE (8321 documentos elegidos aleatoriamente)	PLN	Anotación por un/a investigador/a clínico/a y un farmacéutico/a	Interna Partición (30%/70%)
Iqbal et al. (2015) Reino Unido	ORnC	Base de datos de South London and Maudsley NHS Foundation Trust (2007-2013)	EA extrapiramidales y otros	HCE (12879 pacientes)	PLN	Anotación por dos farmacéuticos/a clínicos/as	Interna (n.r.)
Pandey et al. (2017) Reino Unido	ORnC	Base de datos de South London and Maudsley NHS Foundation Trust	EAM	HCE (6000 casos)	PLN (varios modelos)	Anotación por el equipo investigador	Interna Cruzada (k10)
Pediatría							
Tang et al. (2017) EEUU	ORnC	Pacientes pediátricos hospitalizados y en urgencias (un hospital) (2010-2012)	RAM (29 medicamentos)	HCE (2 647 746 notas de 71 909 encuentros con 42995 pacientes)	PLN	Un/a experto/a en informática clínica y un/a estudiante graduado/a evaluaron la concordancia con los EAM descritos en FAERS	Interna 35.4% de los positivos elegidos aleatoriamente
Otros							

Tabla 7. Características de los estudios primarios incluidos que evaluaron el rendimiento diagnóstico de los modelos desarrollados.

	Diseño	Población	EAM/RAM	Fuente de datos	Modelo	Criterio de referencia	Validación
Chan et al. (2017) Singapur	ORnC	Pacientes tomando estatinas en hospital terciario (2009-2011)	Miopatía por estatinas	HCE (31211 informes; 120 revisados manualmente)	Algoritmo con datos de laboratorio, diagnóstico y prescripción al alta	GS ¹ : Dos fármacos/as independientes, con un/a tercero/a para desacuerdos	Externa M1: 159 positivos revisados manualmente M2: 163 positivos revisados manualmente
Chen et al. (2020) China	ORnC	Pacientes hospitalizados/as con linezolid, vancomicina, gemcitabina, o levofloxacina (2015-2016)	Trombocitopenia Anemia Daño hepático Daño renal	HCE (2692 registros)	Algoritmo con datos de laboratorio	GS ¹ : Cuatro fármacos/as y 2 médicos/as	Externa 145 positivos revisados manualmente

¹Gold standard: verificación de RAM real.

AB-LSTM: Attentive Bidirectional Long Short-Term Memory; BD: Base de datos; DL: Deep learning; DNN: Deep neural network; EAM: Eventos adversos a medicamentos; FAERS: FDA Adverse Event Reporting System; FL: Fuzzy logic; GBA: Gradient boosting algorithm; GS: Gold standard; HCE: Historia clínica electrónica; IPCI: Inhibidores de punto de control inmunitario; LLM: large language model; ML: Machine learning; OPnC: Estudio observacional prospectivo no controlado; ORnC: Estudio observacional retrospectivo no controlado; PLN: Procesamiento del lenguaje natural; RAM: Reacciones adversas a medicamentos; RF: Random forest; TOD: Terapia oral dirigida; TDL: Temporal difference learning; WE: Word embedding.

IV.1.3. Riesgo de sesgo de los estudios incluidos.

En la Tabla 8 se muestra la evaluación completa del riesgo de sesgo para la RS de Murphy et al. (2023) [31]. Se ha valorado como de calidad moderada, dada ciertas limitaciones en la cantidad de bases de datos usadas, reporte de estudios excluidos y evaluación del riesgo de sesgo.

Tabla 8. Calidad metodológica de la revisión incluida (escala AMSTAR-2)	
Dominio	Murphy 2023
1. ¿Las preguntas de investigación y los criterios de inclusión para la revisión incluyen los componentes PICO?	Sí
2. ¿El reporte de la revisión contiene una declaración explícita de que los métodos de la revisión fueron establecidos con anterioridad a su realización y justifica cualquier desviación significativa del protocolo?	Sí
3. ¿Los/as autores/as de la revisión explicaron su decisión sobre los diseños de estudio a incluir en la revisión?	Sí
4. ¿Los/as autores/as de la revisión usaron una estrategia de búsqueda bibliográfica exhaustiva?	Sí parcial
5. ¿Los/as autores/as de la revisión realizaron la selección de estudios por duplicado?	Sí
6. ¿Los/as autores/as de la revisión realizaron la extracción de datos por duplicado?	Sí
7. ¿Los/as autores/as de la revisión proporcionaron una lista de estudios excluidos y justificaron las exclusiones?	No
8. ¿Los/as autores/as de la revisión describieron los estudios incluidos con suficiente detalle?	Sí
9. ¿Los/as autores/as de la revisión usaron una técnica satisfactoria para evaluar el riesgo de sesgo de los estudios individuales incluidos en la revisión?	Sí parcial
10. ¿Los/as autores/as de la revisión reportaron las fuentes de financiación de los estudios incluidos en la revisión?	No
11. Si se realizó un MA, ¿los/as autores/as de la revisión usaron métodos apropiados para la combinación estadística de resultados?	NA

Tabla 8. Calidad metodológica de la revisión incluida (escala AMSTAR-2)	
Dominio	Murphy 2023
12. Si se realizó un MA, ¿los/as autores/as de la revisión evaluaron el impacto potencial del riesgo de sesgo en estudios individuales sobre los resultados del MA u otra síntesis de evidencia?	NA
13. ¿Los/as autores/as de la revisión consideraron el riesgo de sesgo de los estudios individuales al interpretar/discutir los resultados de la revisión?	Sí
14. ¿Los/as autores/as de la revisión proporcionaron una explicación satisfactoria y discutieron cualquier heterogeneidad observada en los resultados de la revisión?	Sí
15. Si se realizó síntesis cuantitativa ¿los/as autores/as de la revisión llevaron a cabo una adecuada investigación del sesgo de publicación (sesgo de estudio pequeño) y discutieron su probable impacto en los resultados de la revisión?	NA
16. ¿Los/as autores/as de la revisión informaron de cualquier fuente potencial de conflicto de intereses, incluyendo cualquier financiamiento recibido para llevar a cabo la revisión?	Sí
CONFIANZA	MODERADA
NA: No aplicable; MA = meta-análisis; RS = revisión sistemática	

En la Tabla 9 se muestra el riesgo de sesgo del EO de efectividad incluido c. El mayor riesgo, como en todos los EO, se debe a la ausencia de aleatorización y el consecuente riesgo de sesgo de selección (equivalencia entre grupos). Además, no se evaluaron posibles factores de confusión.

Tabla 9. Riesgo de sesgo de los estudios observacionales incluidos. Lista de verificación del Joanna Briggs Institute	
Dominio	Smith 2018
1. ¿Fueron los dos grupos similares y se reclutaron de la misma población?	Incierto
2. ¿Se midieron las exposiciones de manera similar para asignar a las personas a grupos expuestos y no expuestos?	Sí
3. ¿Se midió la exposición de manera válida y confiable?	Sí
4. ¿Se identificaron factores de confusión?	No

Tabla 9. Riesgo de sesgo de los estudios observacionales incluidos. Lista de verificación del Joanna Briggs Institute	
Dominio	Smith 2018
5. ¿Se establecieron estrategias para hacer frente a los factores de confusión?	N. A
6. ¿Los grupos/participantes no reportaban la aparición de las medidas de resultado de interés al comienzo del estudio (o en el momento de la exposición)?	N. A
7. ¿Se midieron los resultados de manera válida y confiable?	Sí
8. ¿Se informó el tiempo de seguimiento y fue lo suficientemente largo para que ocurrieran los resultados?	Sí
9. ¿Se completó el seguimiento y, de no ser así, se describieron y exploraron los motivos de la pérdida durante el seguimiento?	Sí
10. ¿Se utilizaron estrategias para abordar el seguimiento incompleto?	N.A.
11. ¿Se utilizó un análisis estadístico apropiado?	Sí
Evaluación global	INCLUIR

IV.1.4. Efectividad y seguridad

IV.1.4.1. Efectividad

Solo un estudio evaluó resultados de efectividad, y sin incluir medidas de resultado finales. Smith et al. (2018) [73] obtuvieron una tasa significativamente menor de prescripción de medicación sospechosa de estar causando RAM (MSRAM) en comparación con el grupo control, durante las primeras 24 horas (28% vs. 39% de la MSRAM detectada, $p < 0.001$; NNT = 9) y durante todo el ingreso (47% vs. 58%, $p < 0.001$; NNT = 9). De la MSRAM prescrita durante el ingreso, el grupo de intervención mostró una tasa menor en las primeras 24 horas (77% vs. 84%, $p = 0.003$; NNT = 8).

No hubo diferencias significativas en la tasa de discontinuación de la MSRAM antes del alta (16% vs 14%, $p = 0.40$).

En la muestra global, no hubo diferencias en la tasa de suspensión/discontinuación de la MSRAM prescrita durante todo el ingreso ($p = 0.28$), pero sí cuando se analizó solo el personal del grupo de intervención que respondió al cuestionario post-alerta, es decir, quienes

definitivamente vieron las alertas (25% vs. 15%, $p = 0.003$; NNT = 8). En cuanto a la MSRAM mencionada en el informe de alta, la diferencia en suspensión/discontinuación no fue significativa en la muestra global (20% vs. 23%, $p = 0.053$), pero de nuevo sí lo fue al incluir solo a quienes respondieron el cuestionario (38% vs. 23%, $p < 0.001$; NNT = 7).

En las respuestas al cuestionario en el grupo de intervención (53% de cuestionarios respondidos), un 74% de las alertas se consideró parcial o totalmente relevante para ser tomada en consideración y en un 57% se consideró posible o seguro que fueron útiles para el manejo del/a paciente. Un 30% de las respuestas declararon que harían o habían hecho cambios en la medicación (confirmados en más de un 90% de dichos casos) mientras que en un 12% no se estaba seguro (de los que un 63% posteriormente realizó cambios).

En una muestra de 100 informes de admisión revisados manualmente, el sistema mostró un VPP de 98% en la detección de la medicación prescrita, 86% en la presencia de RAM en el informe y 97% en la extracción de datos de laboratorio.

El personal médico participante generalmente vio ADER de forma positiva, aunque se sugirió que podría ser más útil para medicamentos que el equipo de medicina internista prescribe con menos frecuencia que los antihipertensivos, como los medicamentos psiquiátricos. Valoraron positivamente la inclusión de posibles causas alternativas de los indicadores clínicos detectados.

IV.1.4.2. Rendimiento diagnóstico

La revisión de Murphy et al. (2023) [31] incluyó 29 estudios publicados entre 2011 y 2021 (17 realizados en EEUU), que desarrollaron y evaluaron sistemas supervisados de PLN para la detección de EAM en la HCE de pacientes hospitalizados. El 48% de los estudios usó datos de sus propias instituciones, mientras que el resto usó bases de datos públicas. Diez estudios (34.5%) no describieron el área clínica de pacientes incluidos; en el resto, las áreas más comunes fueron oncología (31%) y cuidados intensivos (21%). Aproximadamente un tercio de los estudios anotaron sus propias bases de datos, mientras que el resto usó datos ya etiquetados. Los EAM se definieron como entidades o como relaciones medicamento-síntoma, salvo en cinco estudios que evaluaron la presencia de EAM a nivel de frase, documento o paciente.

Un 58.6% de los estudios evaluaron tareas de reconocimiento de entidades, y un 51.7% de extracción de relaciones y clasificación. Solo cuatro estudios evaluaron ambas tareas conjuntamente. Las técnicas de

ML usadas más frecuentemente fueron *Long Short Term Memory* (LSTM) (55.2%) y *Conditional Random Field* (CRF) (37.9%).

Se muestran los resultados del mejor modelo de cada estudio en el rendimiento global, mediante la F1-score. Una mayoría de estudios obtuvo valores por encima de 0.80, y en un 34% por encima de 0.90. Sin embargo, en la detección específica de EAM como entidad o relación, el rendimiento fue mucho más bajo, inferior a 0.53 en 11 de 13 estudios.

En cuanto a los estudios primarios, en la Tabla 10 se muestran los resultados de rendimiento global y para EAM/RAM específicos (en su caso) de los estudios individuales. Para los estudios que evaluaron varios modelos, se muestra el de mejor rendimiento. En el caso de la F-score, se ofrece el valor micro-promediado.

General

Ocho estudios usaron bases de datos con pacientes de diferentes especialidades o estas no son reportadas. Bagattini et al. (2019) [59] desarrollaron un marco de clasificación para transformar series temporales con datos escasos (como es típico de los registros clínicos) en una representación con valor único que puede ser usada por cualquier técnica de ML. En el artículo usan un modelo de RF para la detección de 15 EAM específicos a partir de datos de laboratorio, con un rango de valores de área bajo la curva (AUC) entre 0.59-0.85 (10 de ellos obtuvieron valores por debajo de 0.80).

McMaster et al. (2019) [69] aplicaron ML a datos estructurados (códigos ICD-10, datos de laboratorio, diagnóstico, días de estancia), obteniendo valores de sensibilidad (SN) = 0.75 y AUC = 0.80 para cualquier RAM, y de 0.80 y 0.82, respectivamente, para RAM potencialmente grave o fatal. En un estudio posterior [49] usaron técnicas de PLN, obteniendo valores de SN = 0.89, especificidad (ES) = 0.81 y AUC = 0.82 para la detección de cualquier RAM. En la detección específica del token que define el RAM, la F-score fue de 0.61.

Santiso et al. (2021) [72] usaron tres bases de datos de los hospitales vascos de Basurto y Galdakao. En la validación externa se obtuvieron valores de SN, valor predictivo positivo (VPP) y F-score micro-promediada de 0.813 para el rendimiento global en los tres casos. En la detección específica de RAM como relación medicamento-síntoma (clase positiva) el rendimiento bajó a 0.585, 0.354 y 0.441, respectivamente para las tres métricas. En un trabajo anterior [41] combinaron técnicas de aprendizaje supervisado y no supervisado, obteniendo valores de SN, VPP y F-score de 0.841 en las tres métricas, mientras que en la detección específica de

RAM como relación medicamento-síntoma los valores fueron de SN = 0.77, VPP = 0.55 y F-score = 0.64.

Jeon et al. (2020) [67], aplicando técnicas de PLN a notas de enfermería, obtuvieron un valor de AC = 0.64 para el mejor modelo en la detección de cualquier EAM.

Tang et al. (2019) [75], mediante técnicas de PLN, obtuvieron en la validación externa valores de SN = 0.98, VPP = 0.81 y F-Score = 0.89 para la detección de medicación, y valores de 0.457, 0.641 y 0.534, respectivamente, para el reconocimiento de la relación medicación-evento.

Por último, Zhao et al. (2015) [78], con técnicas de ML aplicadas a datos estructurados, obtuvieron valores promedio de AUC = 0.76 y *accuracy* (AC) = 0.85 para los 27 EAM evaluados. En la detección de EAM individuales, los valores estuvieron entre 0.55-0.98 para el AUC y 0.61-0.99 para AC.

Cáncer

Tres estudios usaron bases con datos de pacientes con cáncer. Van Laar et al. (2020) [76] usaron un modelo de PLN para extraer de la HCE datos demográficos, clínicos y de resultados del tratamiento, incluidos EAM, en pacientes con carcinoma renal metastásico. En el caso de los 4 EAM evaluados, el F1-score estuvo entre 0.75 (diarrea) y 0.50 (síndrome mano-pie).

Ilvalanien et al. (2021) [63] evaluaron prospectivamente la detección de EAM inducidos por los inhibidores de punto de control inmunitario, a partir de síntomas reportados por los pacientes mediante un instrumento digital desarrollado previamente por los autores, analizados con ML. Obtuvieron altos índices de rendimiento en la detección de la presencia de cualquier EAM (AUC: 0.99; F-score: 0.96). Dado el pequeño tamaño muestral, no se pudo evaluar el rendimiento sobre síntomas específicos.

Quillet et al. (2018) [71], en pacientes con cáncer con terapia oral dirigida aplicaron un algoritmo simple mediante filtros aplicados a los códigos ICD-10 registrados en la HCE tres meses después de comenzar la terapia. La mayor ES obtenida (0.98), a partir de la combinación de cuatro filtros, produjo una SN de 0.32.

Reacciones alérgicas

Tres estudios evaluaron la capacidad de sus modelos, todos de PLN, en la detección de reacciones alérgicas a medicamentos. Banerji et al. (2020)

[60] obtuvieron valores de SN = 0.77, ES = 0.89, VPP = 0.86 y F-score = 0.81.

Inglis et al. (2021) [64] tenían como objetivo discriminar entre alergia e intolerancia a la penicilina. El mejor modelo evaluado obtuvo un alto rendimiento, con valores de 0.99 para AUC, ES, VPP, F-score y AC, y de 0.96 para SN. Los resultados fueron igual de buenos en la detección de alergias de alto riesgo.

El modelo de ML de Yang et al. (2020) [77] obtuvo un AUC de 0.98 en la muestra de entrenamiento, y un VPP de 0.25 en el total de las 3 muestras de validación (la segunda con nuevos pacientes del mismo hospital, y la tercera de un hospital distinto).

Psiquiatría

Tres estudios del mismo grupo de investigación usaron bases de datos psiquiátricas para desarrollar sus modelos, todos con PLN. Iqbal et al. (2017) [66] evaluaron la capacidad de detección de 19 EAM asociados a medicación antipsicótica y antidepresiva. Siete EAM alcanzaron un F-score de 0.90 o superior: acatisia (0.93), estreñimiento (0.91), diarrea (0.92), hipersalivación (0.96), náusea (0.92), taquicardia (0.94) y ganancia de peso (0.90), y solo 3 de ellos obtuvieron un valor menor a 0.86. En un trabajo anterior del mismo autor [65] se obtuvieron valores de SN entre 0.86-0.90 y VPP entre 0.85-0.97 en la detección de 4 EA extrapiramidales (disonía, acatisia, parkinsonismo, discinesia tardía), y valores de SN entre 0.69-0.83 y VPP entre 0.64-0.96 para 8 EA no extrapiramidales (alopecia, convulsiones, hipersalivación, miocarditis, náusea, neumonía, síndrome de Stevens-Johnson y taquicardia). Por último, Pandey et al. (2017) [70] obtuvieron valores de SN, VPP Y F-score superiores a 0.80 en la detección de cualquier RAM.

Pediatría

Tang et al. (2017) [74] obtuvieron un VPP de 0.42 en la detección de EAM en una muestra de pacientes pediátricos hospitalizados.

Otros

Dos estudios aplicaron algoritmos con reglas simples a partir de datos de laboratorio, diagnóstico y medicación. Chan et al. (2017) [61] obtuvieron valores de F-score de 0.75 y 0.71 (dos muestras de validación) para detectar miopatía inducida por estatinas. Chen et al. (2020) [62]

obtuvieron un valor de 0.96 en la detección de anemia inducida por gemcitabina, y de 0.44, 0,48 y 0.61 para daño renal por vancomicina, daño hepático por levofloxacino, y trombocitopenia por linezolid, respectivamente.

Tabla 10. Resultados de los estudios primarios que evaluaron el rendimiento diagnóstico de los modelos desarrollados.								
	Modelo	EAM/RAM	AUROC	SN	ES	VPP	F-score	Accuracy
General								
Bagattini (2019)	RF	15 EAM	Rango 0.592-0.848	-	-	-	-	-
McMaster (2019)	RF	Cualquier RAM RAM grave o fatal	0.803 0.822	0.75 0.80	-	-	-	-
McMaster (2023)	MeDeBERTa	Cualquier RAM	Externa 0.916	Externa 0.886	Externa 0.810	-	Interna RRAM: 0.611	
Jeon (2020)	TD-BLR NB	Cualquier EAM	-	-	-	-	-	0.63 0.64
Santiso (2019)	RF	Cualquier RAM	-	Externa Global: 0.841 RME: 0.767	-	Externa Global: 0.841 RME: 0.548	Externa Global: 0.841 RME: 0.639	-
Santiso (2021)	Joint AB-LSTM network	Cualquier RAM	-	Externa Global: 0.813 RME: 0.585	-	Externa Global: 0.813 RME: 0.354	Externa Global: 0.813 RME: 0.441	-
Tang (2019)	FL	Cualquier EAM	-	Interna RM: 0.968 RME: 0.586 Externa RM: 0.979 RME: 0.457	-	Interna RM: 0.949 RME: 0.757 Externa RM: 0.811 RME: 0.641	Interna RME: 0.661 Externa RM: 0.887 RME: 0.534	-
Zhao (2015)	RF	27 EAM	Promedio 0.763 Rango: 0.552-0.980	-	-	-	-	Promedio 0.845 Rango: 0.606-0.989
Cáncer								
Iivanainen (2021)	GBA	Cualquier EAM	0.99	-	-	-	0.94	0.97
van Laar (2020)	-	Diarrea Hepatotoxicidad Hipertensión Síndrome mano-pie	-	0.686 0.872 0-717 0.632	-	0.814 0.493 0.516 0.462	0.745 0.630 0.600 0.533	-

Tabla 10. Resultados de los estudios primarios que evaluaron el rendimiento diagnóstico de los modelos desarrollados.

	Modelo	EAM/RAM	AUROC	SN	ES	VPP	F-score	Accuracy
Quillet (2018)	Combinación de 4 filtros	Cualquier EAM	-	0.324	0.979	-	-	-
Reacciones alérgicas								
Banerji et al. (2020)	MTERMS	Cualquiera	-	0.77	0.89	0.86	0.81	-
Inglis (2020)	ANN	Cualquiera	0.994	0.99	0.96	0.99	0.99	0.99
		Alto riesgo	0.988	0.99	0.99	0.98	0.98	0.99
Yang (2020)	DNN	Cualquiera	0.98 (ME)	-	-	0.25 (M2-4)	-	-
Psiquiatría								
Iqbal (2015)	GATE software	EA extrapiramidal Otros	-	0.86-0.90 0.69-0.83	-	0.85-0.97 0.64-0.96	-	-
Iqbal (2017)	GATE software	19 EAM		Rango 0.68-0.95	Rango 0.62-0.93	Rango 0.51-0.97	Rango 0.61-0.96	Rango 0.72-0.93
Pandey (2017)	RNNACRF	Cualquier	-	0.832		0.822	0.827	
Pediatría								
Tang (2017)	"automated NLP techniques"	Cualquiera	-	-	-	0.424	-	-
Otros								
Chan (2017)	Algoritmo con datos de laboratorio, diagnóstico y medicación	Miopatía por estatinas	M1: 0.781 M2: 0.722	-	M1: 0.714 M2: 0.520	M1: 0.746 M2: 0.605	-	-
Chen (2020)	Algoritmo con datos de laboratorio y medicación	Trombocitopenia Daño renal Daño hepático Anemia	-	-	0.610 0.444 0.479 0.958	-	-	-

AB-LSTM: Attentive Bidirectional Long Short-Term Memory; **ANN:** Artificial neural network; **AUROC:** Área bajo la curva ROC; **EAM:** Eventos adversos de la medicación; **ES:** Especificidad; **GATE:** General Architecture for Text Engineering; **FL:** Fuzzy logic; **GBA:** Gradient boosting algorithm; **DNN:** hierarchical attention-based deep neural network; **M:** Muestra; **ME:** Muestra de entrenamiento; **MTERMS:** Medical Text Extraction, Reasoning, and Mapping System; **NB:** Naïve Bayes; **RAM:** Reacciones adversas a medicamentos; **RM:** Reconocimiento de la medicación; **RME:** Relación medicación-evento; **RNNA-CRF:** Bi-Directional Recurrent Neural Network with Attention with conditional random fields; **RRAM:** Reconocimiento del RAM (token); **SN:** Sensibilidad; **VPP:** Valor predictive positivo.

IV.1.5. Calidad de la evidencia de efectividad y seguridad

Dado que solo se identificó un estudio de efectividad sin medidas finales de resultado, la calidad de la evidencia no se evaluó formalmente mediante el sistema GRADE y se comenta en la discusión.

IV.2. Revisión sistemática de evaluaciones económicas

IV.2.1. Resultado de la búsqueda bibliográfica

La selección de estudios sobre EE parte de la búsqueda bibliográfica realizada para evaluar la efectividad y seguridad de los sistemas automatizados para la detección de RAM.

Los resultados de la búsqueda bibliográfica en las bases de datos electrónicas se pueden consultar en el apartado IV.1.1 del presente informe.

A partir de la lectura de títulos y resúmenes no se seleccionó ninguna referencia que cumpliera con los criterios de selección de coste-efectividad para su lectura a texto completo, por lo que finalmente ningún estudio fue incluido en la RS de EE.

IV.3 Análisis económico

IV.3.1. Estudio de costes

El análisis se basa en los costes de implantación y puesta en marcha de un sistema como el propuesto por la empresa Costaisa descrito en el apartado de metodología del presente informe, en base a la información proporcionada por ellos mismos.

En primer lugar, se requiere un coste inicial de 25 000 € correspondientes a la implantación del sistema y el acompañamiento para realizar la integración del mismo. Para ello, se requiere de un/a farmacéutico/a y un analista programador por parte de la empresa para realizar la implantación, junto con equipos de sistemas y de farmacia de la organización sanitaria.

A continuación, una vez implantado el sistema, es necesario un coste recurrente anual de 45 000 € correspondientes a mantenimiento, actualización de la base de datos y costes propios de la infraestructura en la nube.

IV. 4. Aspectos éticos, legales, organizativos, sociales y ambientales

IV.4.1. Resultados de las consultas con las partes interesadas y expertos/as en informática

En la primera consulta con participantes relevantes celebrada el día 20 de abril de 2023, se identificaron sus principales preocupaciones y expectativas. Los temas relevantes para pacientes, profesionales e industria se recogen en la Tabla 11.

Tabla 11. Resultados de las consultas realizadas a los/as grupos de interés	
Beneficios esperados de la aplicación de sistemas automatizados de detección de RAM	• Posibilidad de detección de riesgo de RAM en tiempo real. • Alta capacidad predictiva en pacientes de alto riesgo de sufrir RAM, como pueden ser pacientes polimedicados/as frágiles y pacientes renales, con neoplasias, pacientes de alto riesgo como inmunodeprimidos/as, población pediátrica. • Identificación de casos con RAM no sospechados por el/la profesional ni el/la paciente. • Posibilidad de identificar características de pacientes que sufren más un determinado tipo de RAM, así como identificar las situaciones en los que las RAM ya conocidas aparecen con mayor frecuencia. • Mejoría de los resultados en salud (morbi-mortalidad y discapacidad) y de la calidad de vida de pacientes y usuarios/as.
Riesgos identificados	• Riesgo de que el sistema tenga un bajo valor predictivo y por tanto genere una sobrecarga de avisos y alertas, así como riesgo de sobre notificación de RAM. • Riesgos relacionados con la confidencialidad de los datos. • Riesgo de que el sistema, si es de inteligencia artificial, sea preciso, pero a costa de ser opaco, ya que los sistemas deben ser transparentes y tener una aplicabilidad aceptable (FAIR principales).

	• Falta de especificidad de los datos, desconocimiento de su manejo o interpretación errónea de los datos proporcionados por el sistema. • Dificultades en la actualización o la demora en la misma compromete la detección temprana de RAM. • Riesgo de que necesite una cantidad excesiva de trabajo técnico para ser eficaz. • Posibilidad de que estos sistemas se centren solo en la cantidad de RAM y no en su calidad (gravedad y novedad). • Falta de cultura institucional de la calidad y gobernanza del dato. • Escasez de equipos interdisciplinarios con formación clínica, técnica en Inteligencia Artificial y científicos de datos en salud. • Posibles sesgos en los algoritmos por raza-etnia, sexo-género u otros determinantes sociales. • Falta de viabilidad en la implantación y desarrollo por necesidad de inversiones iniciales importantes. • Escasa implicación actual de las sociedades científicas en la monitorización y evaluación. • Ambiente de “falsa seguridad” en los/as profesionales asistenciales abandonando buenas prácticas actuales.
Pacientes que potencialmente pueden tener más beneficios	• Población pediátrica que ocasionalmente toman un fármaco. • Pacientes con inmunosupresión, pacientes con procesos oncológicos, que tienen alto riesgo de complicaciones. • Pacientes polimedicados/as, dada su alta prevalencia. • Pacientes con limitaciones cognitivas o de comunicación que no puedan expresar o identificar las RAM que se puedan producir con el tratamiento farmacológico. • Pacientes en la transición del alta hospitalaria a la atención primaria y cuando se añaden medicamentos a lo que ya tomaban es una situación de alto riesgo y que probablemente está muy infra diagnosticado, especialmente en pacientes muy polimedicados/as. • Pacientes con enfermedades psiquiátricas, ya que en psiquiatría el primer mes de tratamiento es crítico para la adherencia, y los efectos adversos es una de las causas más frecuentes de abandono. • Pacientes con enfermedades raras para los que la utilización de fármacos también puede ser un perfil de interés.
Retos organizativos identificados	• Es necesario contar con una supervisión de expertos/as • Evaluar falta de personal para la implementación de estos sistemas • Determinar cuándo y a quién debe llegarle el aviso

- Contemplar la necesidad de formación del personal que vaya a hacer uso del sistema.
- Posibilidad de inconveniencias técnicas como un ciber ataque
- La necesidad de la interoperabilidad para tratar de implantar un sistema único en toda España
- Tener en cuenta el tiempo que consume a la hora de valorar la implantación, puesto que si consume mucho tiempo va a ser difícil implementarlo
- Tener en cuenta el papel del/la paciente como registrador/a, por ejemplo, a través de una App en la Tarjeta Sanitaria Virtual

Además de esta primera consulta con participantes relevantes para poder contextualizar la tecnología y centrar su aplicación al ámbito español, se contactó con cinco expertos/as en informática, obteniendo una única respuesta. En concreto, se llevó a cabo una entrevista con el subdirector TIC en la provincia de Andalucía. Comenzamos señalando los hallazgos obtenidos tras esta consulta, cuyas preguntas pueden ser consultadas en el anexo 6.

IV.4.1.1 Problemas organizativos

Uno de los problemas desde el punto de vista técnico es que los equipos informáticos de los hospitales muchas veces empiezan de cero a la hora de introducir un sistema de detección automatizada de RAM. Partir de un modelo existente en otro hospital es un facilitador siempre que exista interoperabilidad. El proceso que exige la oficina de interoperabilidad hace que la incorporación de estos sistemas sea más lenta. Sin embargo, también constituye una garantía para la seguridad y la calidad del sistema.

IV.4.1.2. Ciberseguridad y calidad de los datos

Particularmente no se identifican desde el punto de vista de los equipos informáticos peligros de ciberseguridad específicos para los sistemas de detección de RAM. El mayor problema parece estar relacionado con la calidad de los datos a partir de los cuales se van a crear los algoritmos que se usarán. En ese sentido, es necesario destacar la importancia de la legibilidad del algoritmo y evitar el funcionamiento tipo “caja negra” donde no se sabe muy bien de qué manera llega el algoritmo a hacer sus predicciones. Para tomar estas decisiones, apoyándose en las predicciones hechas por estos algoritmos, la legibilidad y la transparencia

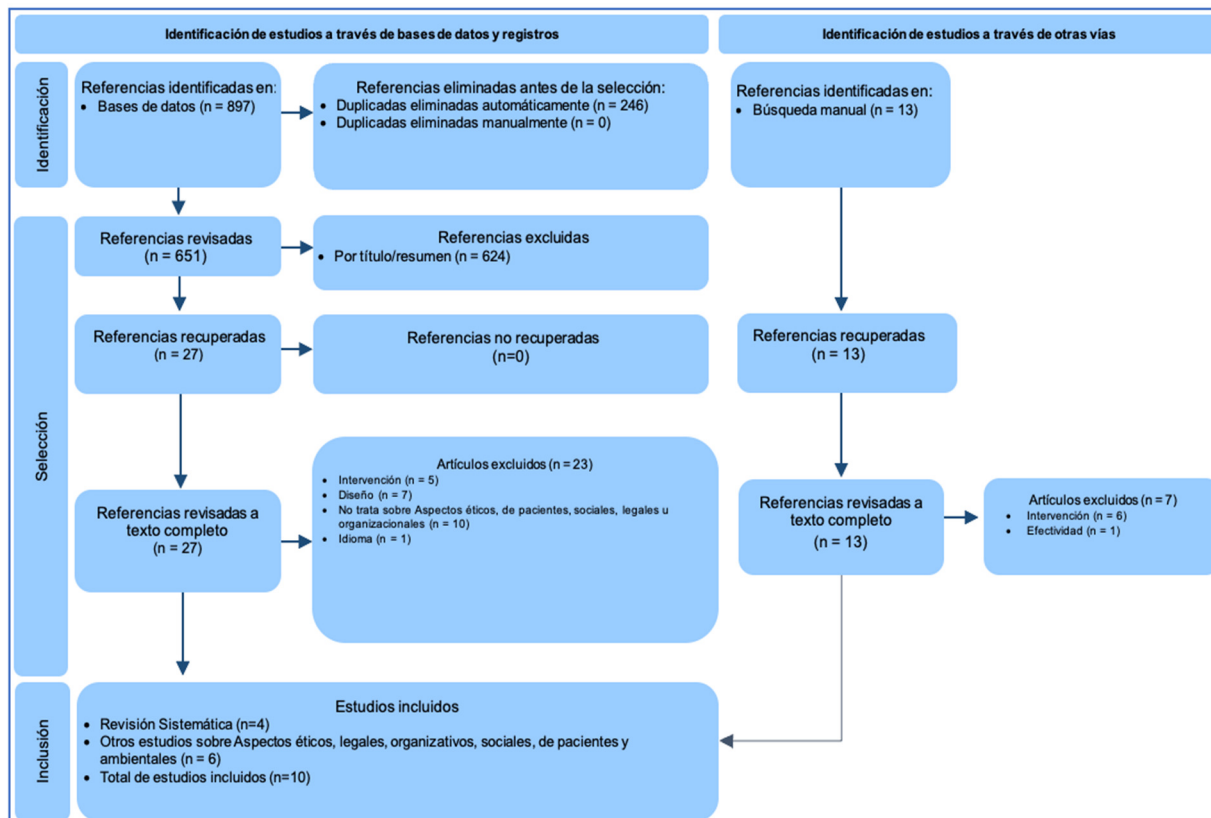
sobre el origen de los datos son altamente deseables. Finalmente, desde el punto de vista de la formación es necesaria al menos una mínima formación tanto para los equipos informáticos como para los equipos clínicos que lo van a usar, especialmente si no hay cultura previa de notificación de RAM.

IV.4.2 Resultados de la Revisión Sistemática

Se localizaron inicialmente 897 referencias bibliográficas, 651 una vez eliminados los duplicados (Tabla 12). Se seleccionaron 27 referencias potencialmente relevantes para ser analizadas en detalle a texto completo, de las cuales 23 fueron excluidas. Finalmente, 4 estudios fueron considerados, a los que se añadieron otros 6 estudios localizados y añadidos tras una búsqueda manual. Un total de 10 estudios se incluyeron en la revisión. El proceso de selección de estudios se muestra en la Figura 2.

Tabla 12. Resultados de la búsqueda bibliográfica (aspectos éticos, legales, organizativos, sociales y ambientales)				
Base de datos	Plataforma de acceso	Periodo buscado	Fecha de acceso	Nº de resultados
MEDLINE	Ovid	1946-2023	02/06/2023	320
Embase	Elsevier	1974-2023	05/06/2023	449
Web of Science Core Collection	Clarivate Analytics	1900-2023	05/06/2023	128
TOTAL				897
Duplicados				246
TOTAL sin duplicados				651

Figura 2. Proceso de selección de estudios de aspectos éticos, legales, organizativos, sociales y ambientales



IV.4.2.1 Características de los estudios incluidos

De los estudios incluidos, cinco son revisiones narrativas [27,79–82], dos estudios de métodos mixtos [28,83], dos estudios retrospectivos mediante revisión de registros [84,85], un estudio descriptivo mediante encuesta [86]. De estos estudios, ninguno fue llevado a cabo en Europa (seis estudios en EEUU, tres estudios en Australia y un estudio en Pakistán). Las principales características de los estudios junto con un resumen de las valoraciones de sus limitaciones metodológicas pueden consultarse en el Anexo 8. La evaluación completa de la calidad metodológica de los estudios puede consultarse en el Anexo 9.

IV.4.2.2. Factibilidad, barreras y facilitadores para la implementación.

El personal sanitario identifica que los sistemas tecnológicos de notificación de RAM integrados en el flujo de trabajo pueden ser facilitadores para la notificación de RAM [28], y que habría más probabilidades de reportar RAM si el proceso fuera mejorado mediante la adopción de una herramienta electrónica automatizada y con la obligatoriedad de la notificación [83] (hay que señalar que en el contexto español la notificación es obligatoria). Las intervenciones futuras destinadas a mejorar la notificación de RAM deben desarrollarse digitalmente para simplificar y automatizar el proceso [27], de modo que no se perciban como una carga administrativa.

A la hora de incorporar sistemas automatizados para la detección de RAM, estos se incorporan, ya sea de manera paralela o integrada, a las HCE. Pero este proceso tiene varias complejidades.

a) Factores que afectan a la factibilidad de sistemas automatizados

En los estudios incluidos, se han identificado varios problemas relacionados con el uso de las HCE para detectar RAM. Estos problemas incluyen la falta de estandarización entre sistemas de HCE, la especificidad en las opciones de ingreso de datos, la documentación incompleta e inexacta, y la fatiga por alerta [79,81,85]. Se destaca la importancia de los registros informales, como notas escritas en papel o en campos estrechos de registros electrónicos, ya que a menudo contienen información sobre RAM que se pasa por alto en la revisión de registros de HCE. El estudio de Hurdle et al. (2003) [85] revela que puede

perderse mucha información relevante para detectar automáticamente RAM si se utilizan exclusivamente los registros de la HCE, ya que se puede perder alrededor del 14 % de la información que se transmite por parte de equipos de enfermería en notas informales, como las notas de los cambios de turno. Sin embargo, al considerar estas notas informales de enfermería, se aumenta la detección de RAM en casi un tercio. Sin embargo, en este estudio, al revisar los datos nuevamente y considerar estas notas informales de enfermería, se aumentó la detección de RAM en casi un tercio [85]. Por tanto, las variaciones en las prácticas de documentación en HCE afectan la capacidad de los sistemas automatizados para detectar RAM, especialmente aquellos basados en PLN, ya que dependen de la consistencia en la información registrada. Las RAM introducidas como texto libre dificultan la estandarización, lo que puede resultar en la pérdida de datos y afectar negativamente la capacidad del sistema automatizado para detectar RAM [79,81,85].

b) Consideraciones para mejorar la implementación

Es importante que los sistemas automatizados para la notificación de RAM estén integrados con la gestión de medicamentos o los registros electrónicos hospitalarios, ya que puede ayudar a ahorrar tiempo y recursos [83]. Sin embargo, la inclusión exclusiva de herramientas electrónicas para la detección automatizada de RAM podría no aumentar la tasa de notificación si no se combinan con estrategias, tales como iniciativas educativas (presenciales o eLearning), incentivos, recordatorios, proporcionar comentarios, entrevistas telefónicas, o mejorar la accesibilidad a los formularios de notificación de RAM [27]. El personal de la salud necesita documentar adecuadamente tanto los tipos específicos de RAM, como la información sobre el momento en que se produjo la reacción, el tiempo transcurrido entre la administración del medicamento, etc. La capacitación sobre notificación de RAM se puede incluir como parte del proceso de incorporación de profesionales de la salud empleados en un entorno hospitalario [83].

IV.4.2.3. Gestión eficaz de las alarmas/notificaciones del sistema.

La fatiga por alertas es un fenómeno que ocurre cuando los/as profesionales quedan inundados con un volumen tan alto de alertas de seguridad que disminuye su capacidad de reaccionar eficazmente, por ejemplo, pudiendo cancelar o anular la prescripción de forma acrítica. Esto puede llevar a insensibilización ante las alertas, lo que pone en

peligro la seguridad [79]. En un estudio sobre las alertas automáticas asociados a algunos medicamentos y los aspectos que determinan su aceptabilidad se encontró que el colectivo médico apoyaba en general el sistema de alertas de medicamentos [86]. Así, el 81% percibía que las alertas de medicamentos mejoraban su conocimiento de las interacciones farmacológicas a la hora de prescribir, y el 71% consideraba que las alertas ayudaban a identificar posibles interacciones farmacológicas, mientras que el 87% consideró que las alertas farmacológicas les ayudaban a aprender sobre las interacciones medicamentosas, y el 79% señaló que las alertas les ayudaban a prescribir con seguridad. En cuanto a los inconvenientes asociados a la exposición a las alertas automáticas de medicamentos, el más reportado fue el exceso de alarmas no relevantes [86]. Así, es necesario contextualizar estos sistemas y comprender las necesidades y preferencias de sus usuarios finales, lo que puede añadir valor en las fases del diseño [28]. En ese sentido, se ha resaltado la importancia crítica de considerar aspectos de la aceptabilidad, usabilidad y experiencia de los/as usuarios/as en el diseño de los sistemas automatizados para detección de RAM [28]. En un estudio realizado por la Universidad de Washington, no se encontraron diferencias en las tasas de anulación de alarmas entre residentes y adjuntos/as ni entre diferentes hospitales [79], aunque estos aspectos deben ser estudiados en futuras investigaciones.

IV.4.2.4. Confidencialidad, ciberseguridad y equidad.

La privacidad es de máxima importancia en los sistemas automatizados de detección de RAM, ya que la información de salud es extremadamente delicada y cualquier filtración de estos datos podría tener graves consecuencias para la vida de las personas. Según la evidencia recopilada, el uso de grandes cantidades de datos (Big Data) ha desempeñado un papel significativo en la identificación de RAM y en el intercambio de información de salud [80]. Se deben explorar nuevas fuentes de datos, desarrollar herramientas de minería de datos y tecnologías de protección de la privacidad [80]. Estos avances permitirán un perfeccionamiento de la farmacovigilancia y, al mismo tiempo, garantizarán la ciberseguridad y la confidencialidad de los datos de salud de las personas.

a) Dilema entre privacidad e intercambio de datos

El manejo de los datos en aplicaciones de salud genera diferentes opiniones sobre qué modelo de consentimiento se debe pedir a los/as

usuarios/as a la hora de compartir sus datos. Así, por un lado, se habla de que se debe emplear un modelo sin consentimiento, en el que el personal sanitario puede utilizar todos los datos médicos de un/a paciente concreto sin su permiso. Frente a ello, otros modelos permiten elegir a los/as pacientes que sus datos estén a disposición de los equipos profesionales sanitarios, de investigación o instituciones de interés. Finalmente, hay modelos más complejos que posibilitan que cada paciente pueda seleccionar el tipo específico de datos que desean revelar y para qué fin [80]. Al compararlos, el modelo en que no se da un consentimiento específico para el uso de los datos tiene más probabilidades de explotar plenamente la información sanitaria para detectar las RAM, a costa de una mayor probabilidad de revelación de datos privados. En los otros dos tipos de consentimiento se protege la privacidad, aunque la preocupación por este aspecto puede dar lugar a una menor disponibilidad de los datos.

Estas condiciones provocan una situación de dilema, que plantea la necesidad de una tecnología avanzada y de políticas jurídicas que satisfagan las dos necesidades: protección de la privacidad y el intercambio de datos [80]. Por otra parte, las instituciones deben centrarse en minimizar las amenazas y las áreas vulnerables que pueden ser atacadas, por ejemplo, mediante autenticación y cifrado de datos [82]. Para minimizar los riesgos asociados a la confidencialidad de los datos, es importante utilizar inicios de sesión seguros que utilicen autenticación multifactor y preguntas de seguridad, tanto para la identificación de pacientes como para la de los/as profesionales [82]. Además, deben realizarse evaluaciones de seguridad periódicas para evitar la divulgación accidental de datos sanitarios y el fraude, así como impartir formación tanto al personal sanitario y usuarios/as para que reconozcan estas amenazas [82].

b) Medidas de seguridad y ciberseguridad

Respecto a la ciberseguridad, los servicios de telemedicina en general suelen requerir conexiones de red inalámbricas que recogen y transmiten información de pacientes a los sistemas de telemedicina, permitiendo a terceras partes acceder a esos datos [82]. Para transmitir y almacenar los datos, los/as usuarios/as dependen de una conexión inalámbrica que puede sufrir múltiples ataques al sistema. El sistema de servicios de telemedicina puede estar expuesto a amenazas de seguridad y ataques de tipo *intermediario* (también conocido como ataque de interposición) [82].

c) Equidad en el acceso a tecnologías, datos y sesgos

Respecto a la equidad, una limitación de los sistemas automatizados de detección de RAM cuando estos dependen de la información que introducen los/as propios/as pacientes, es la dificultad del acceso que algunos/as pacientes pueden tener a servicios de internet, teléfonos inteligentes o dispositivos de monitorización, lo que podría dar lugar a desigualdades en el tratamiento y los resultados de sus datos. Las predicciones imprecisas pueden deberse a lagunas en la información o a la manipulación de los datos clínicos [82].

Finalmente, una cuestión que puede afectar a la equidad es la existencia de sesgos potenciales en los conjuntos de datos utilizados para entrenar los algoritmos de sistemas de IA, ya que estas bases de datos pueden reforzar los sesgos existentes, carecer de datos de poblaciones infrarrepresentadas o no garantizar la equidad de los algoritmos, pudiendo haber además dificultades en torno a la detección de RAM no documentadas y desconocidas, incluidos los eventos raros [82]. El riesgo de sesgo se incrementa cuando se pueden generar algoritmos de “caja negra”, es decir, que no permiten explicar sus predicciones. Otras consideraciones incluyen el mantenimiento de la integridad de los datos, minimizando el uso de conjuntos de datos sesgados y optimizando los datos de HCE, para lo que puede ser importante contar con un departamento informático, pero también con un/a especialista de apoyo en el equipo de telemedicina. Se debe garantizar la optimización de estos modelos antes de integrarlos en la atención rutinaria a los pacientes [82]. Dado que existen muchas preocupaciones relacionadas con la posibilidad de sesgos en los sistemas de IA, pero que escapan al alcance de la RS llevada a cabo, esta cuestión se aborda de manera más amplia en el siguiente apartado.

IV.4.2.5 Síntesis de las sugerencias derivadas de la revisión de los aspectos éticos, legales, organizativos y sociales.

Algunas sugerencias o propuestas para abordar estos desafíos organizativos y mejorar la integración de los sistemas automatizados pueden ser:

- Estandarización de datos: Establecer estándares claros y uniformes para la documentación de RAM en las HCE, como podría ser el uso de MedDRA para describir los RAM de manera consistente.

- Mejora de usabilidad y experiencia de usuario de estos sistemas: Diseño de sistemas automatizados de notificación de RAM con interfaces intuitivas y fáciles de usar (esto puede ayudar a reducir la fatiga por alerta al simplificar el proceso y hacerlo menos intrusivo en el flujo de trabajo clínico).
- Educación y capacitación: Proporcionar formación adecuada a todos los profesionales sobre cómo utilizar eficazmente estos sistemas automatizados. Esto puede incluir sesiones de formación periódica y otro tipo de materiales educativos.
- Seguimiento y retroalimentación: Establecer procesos de seguimiento periódicos para evaluar la calidad de la documentación de las RAM y proporcionar retroalimentación. Identificar áreas de mejora y oportunidades para optimizar el flujo de trabajo puede ayudar a abordar la documentación incompleta o inexacta.
- Colaboración interdisciplinaria: Fomentar la colaboración entre los diferentes profesionales de la salud en el diseño e implementación de los sistemas automatizados.

Con respecto a la privacidad, se debe considerar en cada caso la adecuada anonimización de los datos, la forma que debe tener el consentimiento informado de los/as pacientes a la hora de ceder sus datos para que se desarrollen sistemas de IA, garantizar que los accesos a los datos sean restringidos, promover evaluaciones periódicas para identificar posibles vulnerabilidades y amenazas y actualizar medidas de seguridad, y prestar atención a la formación y concientización de todo el personal implicado respecto de la importancia de proteger la privacidad y el cumplimiento de las normas relacionadas (p.ej. Reglamento general de protección de datos a nivel Europeo, GDPR).

IV.4.2.6 Calidad metodológica de los estudios incluidos

En el Anexo 9 puede verse el detalle de la evaluación de la calidad metodológica de los estudios incluidos, evaluados con los siguientes instrumentos: CASPe [56] para los estudios cualitativos y SANRA [57] para las revisiones narrativas, el instrumento de lectura crítica *Mixed Methods Appraisal Tool* [58] para los estudios de métodos mixtos, la escala de valoración de la evidencia para series de casos del IHE y la lista de comprobación para la evaluación crítica de estudios transversales de JBI (JBI web).

En general, la calidad de los estudios incluidos es entre moderada y alta, salvo para los estudios de series de casos, para los que la calidad es baja. En el caso concreto de las revisiones narrativas [27,79–82] se justifica la importancia del artículo para los/as lectores/as, se especifican los objetivos o preguntas de investigación, se presenta evidencia apropiada apoyada suficientemente por referencias suficientes, y los datos son presentados adecuadamente. Sin embargo, solo dos [27,81] de las cinco revisiones narrativas revisadas incluyeron una descripción de la estrategia de búsqueda.

En el único estudio cualitativo [28] incluido se definió de forma clara los objetivos y eran congruentes con la metodología cualitativa. En este estudio, la metodología y la estrategia de selección de participantes eran adecuadas para alcanzar los objetivos de investigación. No se incluyó ninguna consideración sobre la reflexividad en la investigación, y los aspectos éticos sólo fueron aludidos al mencionar la inclusión de la aprobación del Comité de Ética. Finalmente, el análisis de datos fue suficientemente riguroso y la exposición de los resultados fue clara y aplicable al contexto.

En el estudio basado en métodos mixtos incluido [83] está justificada adecuadamente la utilización de este diseño metodológico para abordar la pregunta de investigación, la integración efectiva de los diferentes componentes del estudio están integrados efectivamente para responder la pregunta de investigación, y es adecuada la interpretación de los resultados de la integración de los componentes cualitativos y cuantitativos. Sin embargo, no se puede evaluar cómo se abordan las divergencias e inconsistencias entre ambos tipos de resultados.

Finalmente, en los dos estudios de series de casos incluidos [84,85] se define claramente el objetivo, metodología y medidas de resultados, pero no se describen las características de las personas participantes y si fueron reclutadas de manera consecutiva, tampoco se describen los criterios de elegibilidad.

IV.4.3 Resultados de la revisión manual de la literatura dirigida

Se llevó a cabo una revisión manual dirigida a entender los sesgos que pueden afectar a los sistemas de IA aplicados al ámbito sanitario, ya que no se encontró en la RS literatura específica referida exclusivamente a los sistemas de IA para la detección de RAM.

En el Anexo 10 se puede encontrar el marco de los principios de la OMS para garantizar una IA responsable (también aplicable a los sistemas

de detección automatizada de RAM que utilicen IA). Si es posible, los algoritmos de IA deberían ser auditados, con especial énfasis en su confiabilidad, trazabilidad y valor ético (Inteligencia Artificial Ética en Sanidad) [87]. A pesar de que no existe actualmente una metodología de “auditoría algorítmica”, ya que se trata de un área nueva, donde existe una experiencia limitada (Inteligencia Artificial Ética en Sanidad) [87], en España se han desarrollado en los últimos años algunos modelos de autoevaluación ética que pueden ser utilizados por entidades públicas o privadas interesadas en diseñar, desarrollar o implementar tecnologías de IA, así como todas las personas que utilicen, compren o reciban estas tecnologías. Estas herramientas pueden ser utilizadas para evaluar los sistemas de detección automatizada de RAM.

- El modelo de autoevaluación PIO [88] (Principios, Indicadores, Observables) se desarrolla desde el Observatorio de Ética en Inteligencia Artificial de Cataluña (OEIAC), que tiene como objetivo principal estudiar las consecuencias éticas, sociales y legales, así como los riesgos y oportunidades de la implantación de la Inteligencia Artificial. Para llevar a cabo el modelo de autoevaluación PIO se tiene que responder de manera afirmativa o negativa a una serie de Indicadores / Observables basados en siete principios éticos considerados fundamentales por diferentes organismos internacionales: (1) la transparencia y explicabilidad, (2) la justicia y equidad, (3) la seguridad y no maleficencia, (4) la responsabilidad y rendición de cuentas, (5) la privacidad, (6) la autonomía, y (7) la sostenibilidad.
- La guía de autoevaluación ética de sistemas de Inteligencia Artificial para aplicaciones de la salud [89] desarrollada en el marco del proyecto “Detección y eliminación de sesgos en algoritmos de triaje y localización para la COVID-19” (Fundación BBVA) es un modelo de verificación en forma de *checklist* que permite autoevaluar el compromiso con la detección y eliminación de sesgos éticos en los algoritmos utilizados en el campo de la salud. El objetivo de esta herramienta es que sirva para tomar conciencia de la importancia de prevenir, detectar y eliminar sesgos que vuelven injustos tanto el diseño como la aplicación de los algoritmos usados en el campo de la salud y, de este modo, contribuir al desarrollo de otros más justos.

IV.5. Aspectos ambientales

En la fase de definición de alcance no se identificó ningún estudio referido a los impactos ambientales de los sistemas automatizados de detección de RAM. En la revisión manual de la literatura a partir de términos de interés ambiental realizada posteriormente (31/10/2023), se identificó una revisión de alcance [90] que evalúa la huella de carbono de intervenciones de salud digital, que incluye un estudio que utiliza la herramienta *Energy Efficient Tester* (EET) [91], y está basado en el marco de evaluación *Framework for Energy Efficiency Testing to Improve eNvironmental Goals of the Software* (FEETINGS). Este estudio desarrollado por universidades españolas, introduce dos medidas de resultado que podrían ser de interés para la tecnología evaluada: consumo de energía por componentes del ordenador y consumo de energía de los registros personales electrónicos (del inglés, *power consumption by PC components and power consumptions of electronic personal records*), que muestran la reciente preocupación por los aspectos vinculados con la eficiencia energética en el desarrollo de software y con la sostenibilidad ambiental en los registros personales de salud, que deberían ser tenidos en cuenta por los desarrolladores de sistemas automatizados de detección de RAM en nuestro contexto.

A la fecha de finalización del informe (19/12/2023), no se obtuvo respuesta por parte de la industria, por lo tanto, no se dispone de información específica de los impactos ambientales de los sistemas automatizados de detección de RAM.

IV.6. Necesidades de investigación e identificación de medidas de resultado

A continuación, se enumeran las diferentes necesidades de investigación y conjuntos de medidas de resultado estándares identificadas a través de las diferentes fuentes de información consultadas. Las necesidades están clasificadas según su temática y fuente de información. Las principales características de los estudios identificados se muestran en el Anexo 11.

IV.6.1 Necesidades de investigación identificadas en la RS realizada para el presente informe

Dada la ausencia de estudios que evaluaran variables finales de efectividad y seguridad, las necesidades de investigación están completamente por cubrir. Idealmente, es necesaria la realización de ECA por conglomerados (aleatorización de centros o unidades clínicas equivalentes), tanto a nivel hospitalario como en atención primaria, que evalúen como mínimo el número de RAM reales evitadas en un seguimiento de al menos seis meses, además de hospitalizaciones, visitas a urgencias (en pacientes no ingresados) y morbilidad/mortalidad.

Las principales necesidades de investigación identificadas referidas a los aspectos éticos, legales, organizativos, sociales, de pacientes y ambientales son las siguientes:

- Es necesario desarrollar estudios que incluyan codiseño de los sistemas automatizados para la detección de RAM. El uso de esta metodología puede identificar necesidades y preferencias del personal sanitario, así como ofrecer soluciones que signifiquen un beneficio real.
- Es necesario conducir estudios que permitan valorar si se puede llevar a cabo una implementación progresiva en situaciones especiales, como la transición del alta hospitalaria al domicilio, de modo que además se permita identificar casos de RAM no sospechados ni por profesionales ni por pacientes.
- Es necesario llevar a cabo estudios que valoren cómo gestionar las notificaciones de manera eficiente para evitar saturación de alarmas en la historia clínica, desde la perspectiva de todos/as los/as profesionales involucrados/as.
- Es necesario validar el uso de herramientas de evaluación ética de los sistemas de inteligencia artificial aplicados a salud en los sistemas de detección automatizada de RAM que utilicen IA.
- Estudios que evalúen la multidisciplinariedad en la detección precoz de RAM mediante sistemas automatizados.

- Es necesario llevar a cabo estudios cualitativos que exploren las experiencias de profesionales con sistemas automatizados de detección de RAM.

IV.6.2 Necesidades de investigación identificadas en estudios específicos de necesidades de investigación

En la base de datos JLA no se identificaron estudios sobre necesidades de investigación para la tecnología de interés de este informe.

IV.6.3 Medidas de resultados estándares

En la base de datos COMET, se encontraron dos estudios que están en marcha y no tienen aún resultados publicados. El primero de ellos tiene como objetivo desarrollar un conjunto de resultados básicos para la seguridad del paciente en la atención perioperatoria, destacando la importancia de los datos de seguridad del paciente para mejorar la calidad de la atención y la toma de decisiones sanitarias [92]. El segundo estudio, ha comenzado en abril de 2023, y pretende desarrollar un conjunto de resultados básicos para la evaluación de un sistema de Entrada de Orden Médica Computarizada (CPOE) y su implementación en la atención hospitalaria [93].

En la base de datos ICHOM no se encontró ningún estudio.

IV.7 Estudios en marcha

Tan solo se identificó un estudio de intervención en marcha sobre detección automatizada de eventos/reacciones adversas en <https://clinicaltrials.gov/>. Es un ECA por conglomerados realizado en Francia (NCT05923983) que evaluará la efectividad de un sistema de alerta de detección de fallo renal agudo o hiperpotasemia inducidos por medicación en pacientes hospitalizados mayores de 65 años. Se evaluarán el número de efectos adversos detectados, los cambios terapéuticos y en el flujo de trabajo, relevancia de alertas y coste-efectividad. La fecha estimada de finalización es julio de 2024.

V. Discusión

V.1 Efectividad y seguridad

La revisión solo identificó un estudio de efectividad (prospectivo con control histórico), que evaluó un sistema de PLN durante 3 meses [73]. Sus resultados mostraron en el grupo de intervención una menor tasa de prescripción de medicación detectada por el sistema como sospechosa de estar causando RAM (el estudio se centró en medicación antihipertensiva), tanto en las primeras 24 horas como durante todo el ingreso. Sin embargo, no hubo diferencias en la tasa de discontinuación de la medicación sospechosa antes del alta. En el momento del alta, esta tasa de discontinuación fue mayor en el grupo de intervención, pero solo cuando se analizaron únicamente los equipos médicos que respondieron al cuestionario post-alarma. Los resultados muestran por tanto que la intervención produjo cambios en la conducta de prescripción del personal médico, aunque también abren otros interrogantes sobre la ausencia de diferencias en discontinuación durante el ingreso o por qué en algunos casos se volvió a prescribir la medicación sospechosa, cuestiones que no pueden ser respondidas sin la evaluación de las RAM ocurridas. Por tanto, se trata de evidencia de calidad muy limitada; en cuanto a su validez interna, se trata de un estudio no aleatorizado, con un período de intervención de solo tres meses, y en el que no se recogió el número de RAM reales ocurridas, ni los resultados de la intervención sobre variables finales de salud. En cuanto a su validez externa, la generalización de los resultados se ve limitada por restringirse a pacientes hospitalizados/as, en un solo centro, y al focalizarse en RAM producidos por una medicación específica y ya eventualmente presentes en el momento del ingreso. En cualquier caso, el resultado sobre la diferencia en la tasa de discontinuación de medicación al alta señala la importancia de una adecuada atención y correcta interpretación de las alarmas por parte de los profesionales, para producir modificaciones efectivas de las pautas de prescripción. Este proceso debe implementarse de la forma menos disruptiva posible en el flujo de trabajo de los profesionales.

En cuanto a los estudios de rendimiento diagnóstico, los resultados de la revisión incluida sobre PLN y de los estudios primarios que aplicaron dicha técnica (ninguno de ellos incluidos en la revisión) muestran tendencias similares. El rendimiento global de los modelos para detectar cualquier RAM es en general bueno, pero cae considerablemente cuando se trata de detectar específicamente los EAM/RAM anotados como

entidad o relación entre entidades (dato reportado solo en cuatro estudios). Esto es así porque los términos que se refieren a medicamentos son inequívocos en la mayoría de los casos, pero los términos referidos a síntomas, aparte de una mayor variación en la forma en que se describen en la HCE (p.ej., somnolencia, sopor), pueden ser RAM, pero también síntomas producidos por la enfermedad. Por tanto, el rendimiento global da una impresión inflada de la capacidad de detección de las RAM. Por otra parte, cuando se trata de detectar RAM específicos, el rendimiento global varía considerablemente entre ellos dentro de un mismo estudio.

Aparte de los resultados comentados, existen varias limitaciones metodológicas de la evidencia obtenida sobre el PLN. Una mayoría de estudios no realizó una validación externa del modelo, y solo en tres de ellos se verificó la ocurrencia real de RAM. Las métricas reportadas varían entre los estudios, y varios de ellos solo ofrecen resultados para una o dos de ellas, ofreciendo así una visión limitada del rendimiento.

En resumen, la investigación sobre el PLN en la detección de RAM a partir de textos narrativos de las HCE se encuentra mayoritariamente en una fase de carácter aún básico, evaluando soluciones para los retos existentes en dicha área como el desequilibrio en las HCE entre número de RAM (en general poco frecuentes) y eventos no-RAM, detección de negaciones en los textos, representación de series temporales, creación de *corpora* de datos anotados en idiomas distintos del inglés, etc.

Los estudios que usaron únicamente datos estructurados (umbrales en los valores de laboratorio, filtros en códigos ICD-10, datos clínicos y sociodemográficos), fueron en su mayor parte los que verificaron la ocurrencia de RAM reales. Entre ellos, los tres que usaron algoritmos simples (umbrales en los valores de laboratorio, filtros en los códigos ICD-10) [61,62,71] obtuvieron un rendimiento más bajo que el mostrado por los estudios que aplicaron ML. Entre estos últimos, Ilvananien et al. (2021) [63] fue el único estudio que uso datos de síntomas reportados por los pacientes (con cáncer) mediante un cuestionario online de 18 ítems desarrollado específicamente para los pacientes objetivo, obteniendo un excelente rendimiento en una pequeña muestra de pacientes. Si bien este procedimiento esquivas las complejidades del PLN en los textos narrativos sin perder riqueza informativa, requiere el cumplimiento periódico de cuestionarios por parte del grupo de pacientes, lo cual conlleva sus propias dificultades de implementación y cumplimiento.

V.2 Coste-efectividad

No se han identificado estudios que evalúen el coste-efectividad de los sistemas automatizados para la detección de RAM que cumplan con los criterios de selección de la presente revisión. Debido a la falta de evidencia, así como a la escasez de información, no se ha podido llevar a cabo una EE completa contextualizada en España. Por lo tanto, se realizó una evaluación económica parcial mediante un análisis del coste de un sistema para interceptar el proceso de prescripción y poder realizar el soporte a prescripción farmacéutica desde la perspectiva del SNS, es decir, teniendo en cuenta únicamente costes directos sanitarios. Debido a la limitada información disponible, este análisis se basó exclusivamente en la información proporcionada por parte de una única empresa respecto de un sistema automatizado actualmente en fase de pilotaje. Se tuvieron en cuenta solo los costes de implantación y acompañamiento para realizar su integración en el Sistema de Información Hospitalaria, estimado en 25 000 €. A este coste habría que añadir un coste recurrente anual de 45 000 € para mantenimiento, actualización de la base de datos y los costes propios de la infraestructura en la nube. Hay que tener en cuenta que este coste hace referencia a un sistema automatizado en particular y no necesariamente puede o deber ser generalizable a cualquier sistema automatizado de detección de RAM.

V.3 Aspectos éticos, legales, organizativos, sociales y ambientales

Siguiendo los temas identificados como relevantes por las partes interesadas en esta tecnología (pacientes, profesionales e industria), los hallazgos de la RS realizada para analizar estos aspectos se han agrupado en los siguientes temas: factibilidad, barreras y facilitadores para la implementación y aceptabilidad; gestión eficaz de las alarmas o notificaciones del sistema; confidencialidad, ciberseguridad y equidad.

Esta RS muestra que las tasas de notificación de RAM pueden mejorar mediante una educación y una cultura positiva de notificación [81,83,84]. Concretamente, desde el punto de vista organizativo, la incorporación de sistemas tecnológicos de notificación de RAM integrados en el flujo de trabajo y la multidisciplinariedad son dos aspectos importantes a considerar, ya que los programas de notificación de RAM en los que

participa el equipo médico, enfermero y farmacéutico han demostrado ser más fructíferos en la gestión y notificación en comparación con aquellos en los que sólo participan equipos médicos [28,81,83]. En España, el factor más importante que determina la notificación de las RAM es el conocimiento sobre el acto de la notificación y sus implicaciones por parte del personal sanitario [94], por lo que no hay que obviar la formación también cuando se incorporen sistemas de detección automatizada, ya que las estrategias combinadas parecen mejorar las tasas de notificación [27].

La planificación adecuada de la implementación de un sistema de detección precoz de RAM es un aspecto crucial para su introducción en el sistema sanitario. Dentro del contexto del SNS, es altamente recomendable incorporar los sistemas automatizados de detección de RAM a través de la oficina de interoperabilidad. A pesar de que este proceso puede requerir más tiempo cuando se incorpora un sistema por primera vez, ayuda a que el tiempo de implantación posterior en otros hospitales sea menor. Además, sirve como un sistema de garantía de calidad y seguridad, ya que se deben cumplir todas las normas de interoperabilidad. Adicionalmente, el uso de metodologías como el codiseño, a la hora del desarrollo de estos sistemas automatizados, puede ser una garantía que estos sistemas, van a responder a las necesidades expresadas por todas las partes interesadas.

Finalmente, cabe recordar que la función principal de los algoritmos de IA utilizados en sistemas de detección precoz de RAM es determinar la probabilidad de que una persona esté desarrollando una RAM. Sin embargo, en ese proceso, estos algoritmos pueden tener dos tipos de errores [95,96]:

Errores de juicio: las estimaciones automáticas pueden ser incorrectas por varios motivos (como el “sobreajuste” de los datos utilizados para entrenar el modelo). Por tanto, podemos tener un falso positivo (si el paciente no está experimentando una RAM, pero el algoritmo se la diagnostica) o un falso negativo (si el algoritmo afirma que la persona no tiene una RAM, pero en realidad la tiene). Estos errores de juicio pueden conducir a intervenciones inútiles que afectan negativamente a la salud física y mental de una persona.

Error de diagnóstico: estos tipos de falsos positivos o falsos negativos son causados por interpretaciones de profesionales de la salud, en lugar de las estimaciones automáticas de un algoritmo. Estos resultados erróneos pueden afectar la salud tanto de pacientes como de los/as

profesionales sanitarios, ya que pueden interferir con el trabajo de diagnóstico y provocar “fatiga de alerta”. Por su importancia, se necesita revisar o cuestionar los resultados producidos por el algoritmo, por lo que es necesario complementarlos con la participación de profesionales en procesos de auditoría algorítmica.

Por ello, es importante resaltar la necesidad de validar los juicios predictivos arrojados por el sistema de detección automatizada de RAM por parte de un/a profesional. Además, los algoritmos de IA aprenden a tomar decisiones en función de los datos de entrenamiento de decisiones humanas y estas pueden ser sesgadas o reflejar desigualdades históricas o sociales. Es necesario estar alerta ante la posibilidad de sesgos en los sistemas de IA [95,96]. Por ello, para minimizar los posibles sesgos en los conjuntos de datos utilizados para entrenar los algoritmos de sistemas de IA en salud, es importante no solo la transparencia a la hora de reportar la procedencia de los datos, sino además incluir expertos/as en ética y desigualdades en los equipos interdisciplinarios encargados de diseñar, validar y trasladar a la clínica los algoritmos de IA.

Para minimizar estos posibles sesgos, además, se han identificado al menos dos herramientas disponibles online, el modelo de autoevaluación PIO (<https://oeiac.cat/es/modelo-de-autoevaluacion-pio/>) y la guía de autoevaluación ética de sistemas de Inteligencia Artificial para aplicaciones de la salud (<https://airesearch-sesgos-covid19.com/>) que permiten tanto a desarrolladores/as de estos sistemas como a los/as usuarios finales hacer una autoevaluación ética para valorar si el sistema en cuestión puede tener altas probabilidades de contener sesgos que puedan comprometer la equidad.

Un aspecto muy importante que impacta en la aparición y la detección de las RAM tiene que ver con los determinantes sociales de la salud. En este sentido, a pesar de que en nuestra revisión no se encontraron estudios que analizaran la relación o el impacto que puedan tener los determinantes sociales de la salud en los sistemas de detección automatizada de RAM, es importante señalar que este elemento debe tenerse en cuenta a la hora de diseñar estos sistemas automatizados.

Es importante enfatizar el papel fundamental de la implicación y formación transdisciplinaria en la detección de RAM, destacando a todos los profesionales sanitarios de los diferentes ámbitos y niveles asistenciales. Es igualmente importante incluir formación en la prevención, detección y abordaje de las RAMS en el ámbito de los grados, post-grados y formación sanitaria especializada.

V.4 Estimación de la fecha de actualización del informe

Si bien solo se ha identificado un estudio en marcha sobre efectividad de un sistema automatizado de detección de RAM, la rápida evolución de la IA y sus aplicaciones recomiendan una actualización del informe en un plazo no mayor a tres años (inicio de 2027).

VI. Conclusiones

- No se ha identificado ningún estudio que evalúe la efectividad y seguridad de los sistemas automatizados de detección de RAM desarrollados en los últimos 10 años en la reducción de RAM y sus consecuencias sobre variables de salud finales. Solo un estudio evaluó la implementación hospitalaria de un sistema de PLN y su efecto sobre la prescripción de medicación antihipertensiva durante 3 meses, mostrando una reducción de la prescripción de medicación evaluada por el sistema como sospechosa de estar produciendo RAM previamente al ingreso.
- Se ha llevado a cabo un buen número de estudios que evalúan el rendimiento diagnóstico de los métodos de IA aplicados a los datos de HCE, especialmente PLN. Estos estudios muestran en general un buen rendimiento global de los modelos, pero que es debido principalmente a la detección de medicación, mientras que los resultados aún están por debajo de lo aceptable en la detección específica de RAM en los textos clínicos. Por otra parte, el rendimiento global muestra una amplia variabilidad cuando se estudian RAM específicas dentro de un mismo estudio. La mayoría de estos estudios solo realizaron una validación interna de sus modelos y no verificaron la ocurrencia de RAM reales más allá de corroborar su mención en los textos para entrenar los modelos.
- No se identificó evidencia sobre el coste-efectividad de los sistemas automatizados para la detección de RAM.
- El coste estimado de un sistema automatizado para interceptar el proceso de prescripción y poder realizar el soporte a prescripción farmacéutica fue de 25 000 €, en concepto de implantación y acompañamiento para realizar su integración en el Sistema de Información Hospitalaria. Para el mantenimiento, actualización de la base de datos y los costes propios de la infraestructura en la nube, habría que añadir un coste recurrente anual de 45 000 €.
- La implementación de un sistema automatizado de detección de RAM debería llevarse a cabo únicamente tras una meticulosa planificación, en conjunción con los profesionales clínicos implicados, que tenga en cuenta el área clínica donde se maximice su efectividad y eficiencia (por ejemplo, en poblaciones

con alto riesgo o en el caso de RAM graves), las expectativas y preocupaciones de los profesionales y el efecto sobre su flujo de trabajo, así como las cuestiones técnicas referidas a la viabilidad de su implementación, su actualización en el futuro, así como su escalabilidad y generalización a otras áreas clínicas y otros centros del sistema de salud.

- Dentro del contexto del SNS, es altamente recomendable incorporar los sistemas automatizados de detección de RAM a través de la oficina de interoperabilidad, que garantiza la calidad y seguridad del sistema, además de facilitar la implementación del mismo sistema en diferentes centros sanitarios. Adicionalmente, el uso de metodologías como el codiseño a la hora del desarrollo de estos sistemas automatizados puede responder a las necesidades expresadas por todas las partes interesadas.

Contribución de los autores y revisores externos

Autores

- *Amado Rivero-Santana* ✉. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Coordinador del informe. Selección de estudios, extracción de datos, lectura crítica de los estudios incluidos, redacción del presente informe.
- *Vanesa Ramos-García*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Gestora del informe. Selección de estudios, extracción de datos, lectura crítica de los estudios incluidos, redacción parcial del presente informe, revisión interna del borrador del informe y de la versión final del informe.
- *Janet Delgado-Rodríguez*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Departamento de Filosofía I, Universidad de Granada. Evaluación de los aspectos éticos, legales, organizativos sociales y ambientales, redacción parcial del informe, revisión interna del borrador del informe y de la versión final del informe.
- *Alezandra Torres-Castaño*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Revisión interna del borrador del informe y de la versión final del informe.
- *Cristina Valcárcel-Nazco*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Revisión de la literatura económica, estudio de costes, redacción parcial del informe, revisión interna del borrador del informe y de la versión final del informe.

- *Himar González-Pacheco*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Identificación de necesidades de investigación y medidas de resultado estándares, redacción parcial del informe, revisión interna del borrador del informe y de la versión final del informe.
- *Analía Abt Sacks*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Evaluación de los aspectos éticos, legales, organizativos, sociales y ambientales redacción parcial del informe, revisión interna del borrador del informe y de la versión final del informe.
- *Ana Toledo-Chávarri*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Evaluación de los aspectos éticos, legales, organizativos, sociales y ambientales, redacción parcial del informe, revisión interna del borrador del informe y de la versión final del informe.
- *Carmen Guirado Fuentes*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Revisión de la literatura económica, estudio de costes, redacción parcial del informe, revisión interna del borrador del informe y de la versión final del informe.
- *Estefanía Herrera-Ramos*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS). Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Diseño y ejecución de las estrategias de búsqueda asociadas a este informe, selección de estudios y revisión interna del borrador del informe y de la versión final del informe.
- *Violeta Cazaña-Pérez*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Selección de estudios, extracción de datos, lectura crítica de los estudios incluidos revisión interna del borrador del informe y de la versión final del informe.

- *Miguel Ángel García-Bello*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Selección de estudios, lectura crítica de los estudios incluidos revisión interna del borrador del informe y de la versión final del informe.
- *Anthea Santos-Álvarez*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Selección de estudios, extracción de datos, lectura crítica de los estudios incluidos revisión interna del borrador del informe y de la versión final del informe.
- *Andrea Duarte-Díaz*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Revisión interna del borrador del informe y de la versión final del informe.
- *Yolanda Álvarez-Pérez* Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Selección de estudios, revisión interna del borrador del informe y de la versión final del informe.
- *Yolanda Ramallo-Fariña*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS), Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). Identificación de Necesidades de Investigación y medidas de resultado estándares, revisión interna del borrador del informe y de la versión final del informe.
- *Paula Masiero Aparicio*. Médico especialista en Farmacología Clínica/Medicina estética e integrativa. Revisión del protocolo y la versión final del informe.
- *Carlos Boada del Campo*. Médico de familia. Centro de Farmacovigilancia de Canarias/Servicio de Farmacología Clínica Del HUC. Revisión del protocolo y la versión final del informe.
- *Juan Antonio López Rodríguez*. Médico de familia. Centro Salud General Ricardos. Unidad Investigación at primaria Madrid. Gerencia asistencial Atención primaria Madrid. Universidad Rey Juan Carlos. Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Revisión del protocolo y la versión final del informe.

- *Laia Domingo Torrell*. Farmacéutica. Hospital del Mar. Barcelona. Revisión del protocolo y la versión final del informe.
- *Cristina Lugones Sánchez*. Enfermera. Unidad de Investigación de Atención Primaria de Salamanca (APISAL). Instituto de Investigación Biomédica de Salamanca (IBSAL). Gerencia de Atención Primaria de Salamanca (SACyL). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Revisión del protocolo y la versión final del informe.
- *Lilisbeth Perestelo-Pérez*. Servicio de Evaluación del Servicio Canario de la Salud (SESCS). Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS). Supervisión del estudio, revisión interna del borrador del informe y de la versión final del informe.



amado.riverosantana@sescs.es

Revisores externos

- *Alicia Begoña Capilla Lanagrán*. Presidenta de la Asociación de Víctimas de la Vacuna del Papiloma (AAVP). Revisión de la versión final del informe.
- *Antonio Manuel Ojeda Cruz*. Jefe de Servicio de Electromedicina. Complejo Hospitalario Universitario Materno Insular. Revisión de la versión final del informe.
- *Berta Candia Bouso*. Vocal de presidenta de la Sociedad Española de Calidad Asistencial. Revisión de la versión final del informe.
- *Cristina Fernández*. Agencia Española de Medicamentos y Productos Sanitarios. Revisión de la versión final del informe.
- *Eduardo López Briz*. Sociedad Española de Farmacia Hospitalaria. Revisión de la versión final del informe.
- *Fernando Vázquez Valdés*. Secretario de presidenta de la Sociedad Española de Calidad Asistencial (SECA). Revisión de la versión final del informe.
- *Gloria Julia Nazco Casariego*. Jefa de servicio de farmacia del HUC. Revisión de la versión final del informe.
- *Javier Plaza Zamora*. Responsable de área de la Sociedad Española de Farmacia Clínica, Familiar y Comunitaria (SEFAC). Revisión de la versión final del informe.

- *Luis Santiago Sánchez Fernández*. Equipo provincial TIC de Sevilla. Subdirección de Tecnologías de la Información y Comunicaciones. Revisión de la versión final del informe.
- *Silvia López de Alaiz*. Responsables de área de la Sociedad Española de Farmacia Clínica, Familiar y Comunitaria (SEFAC). Revisión de la versión final del informe.

Representantes de asociaciones de pacientes y sociedades científicas

Representantes de asociaciones de pacientes

- *Bernard Gaspar Martínez*. Asociación Española de afectados de cáncer de pulmón (AECAP). Revisión del protocolo del informe.
- *Juan Francisco Perán*. Federación Española de Diabetes (FEDE). Revisión del protocolo del informe y la versión final del informe.
- *Marta Moreno Barón*. Federación Nacional de Asociaciones para la Lucha de las enfermedades del riñón (ALCER). Revisión del protocolo del informe.

Representantes de sociedades científicas

- *Alberto Zamora Cervantes*. Sociedad Española de Medicina Interna (SEMI). Revisión del protocolo del informe y la versión final del informe.
- *Antonio Villafaina Barroso*. Asociación Para las Prácticas Innovadoras en Polimedicación y Salud, integrada en la Fundación para la Innovación, Investigación, Formación y el Desarrollo Comunitario (FÜNDEC). Revisión del protocolo del informe y la versión final del informe.
- *Eduardo López Briz*. Sociedad Española de Farmacia Hospitalaria. Revisión de la versión final del informe.
-
- *Francisco Ferre Navarrete*. Sociedad Española de Psiquiatría y Salud Mental (SEPSM). Revisión del protocolo del informe y la versión final del informe.
- *Gemma Rodríguez Palomar*. Sociedad Española de Farmacéuticos de Atención Primaria (SEFAP). Revisión del protocolo del informe y la versión final del informe.

- *Joaquín Saez Peñataro*. Sociedad Española de Farmacología Clínica (SEFC). Revisión del protocolo del informe y la versión final del informe.
- *Miguel Ángel Hernández Rodríguez*. Sociedad Española de Medicina de Familia y Comunitaria (semFYC). Revisión del protocolo del informe y la versión final del informe.

Los comentarios de los representantes de pacientes y sociedades científicas se muestran en el Anexo 12.

Declaración de intereses

El equipo autor del presente informe y sus revisores externos/as completaron un formulario de declaración de intereses.

El equipo autor del presente informe declara no tener conflictos de intereses en relación con la tecnología evaluada y los comparadores considerados.

Referencias

1. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Care Services, Board on Global Health, Committee on Improving the Quality of Health Care Globally. Crossing the Global Quality Chasm: Improving Health Care Worldwide [Internet]. Washington (DC): National Academies Press (US); 2018 [citado 13 de marzo de 2023]. (The National Academies Collection: Reports funded by National Institutes of Health). Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK535653/>
2. Seguridad del paciente [Internet]. [citado 13 de abril de 2023]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/patient-safety>
3. Agencia Española de medicamentos y productos sanitarios (AEMPS). Información para las notificaciones de sospechas de reacciones adversas a medicamentos por parte de profesionales sanitarios [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2019 [citado 13 de marzo de 2023]. Disponible en: <https://www.aemps.gob.es/medicamentos-de-uso-humano/farmacovigilancia-de-medicamentos-de-uso-humano/notificacion-de-sospechas-de-reacciones-adversas-a-medicamentos-ram-de-uso-humano/notificasospechas-ram-profsanitarios/>
4. Agencia Española de medicamentos y productos sanitarios (AEMPS). ¿Qué es el Sistema Español de Farmacovigilancia de medicamentos de uso humano? [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. 2019 [citado 13 de marzo de 2023]. Disponible en: <https://www.aemps.gob.es/medicamentos-de-uso-humano/farmacovigilancia-de-medicamentos-de-uso-humano/que-es-el-sistema-espanol-de-farmacovigilancia-de-medicamentos-de-uso-humano/>
5. World Health Organization. La farmacovigilancia: garantía de seguridad en el uso de los medicamentos [Internet]. Organización Mundial de la Salud; 2004 [citado 13 de marzo de 2023]. Report No.: WHO/EDM/2004.8. Disponible en: <https://apps.who.int/iris/handle/10665/68862>
6. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet Lond Engl. 7 de octubre de 2000;356(9237):1255-9.

7. Barsky AJ. Nonspecific Medication Side Effects and the Nocebo Phenomenon. *JAMA*. 6 de febrero de 2002;287(5):622.
8. European Comission. Strengthening pharmacovigilance to reduce adverse effects of medicines [Internet]. European Commission - European Commission. [citado 13 de marzo de 2023]. Disponible en: https://ec.europa.eu/commission/presscorner/detail/de/MEMO_08_782
9. Allué N, Chiarello P, Bernal Delgado E, Castells X, Giraldo P, Martínez N, et al. Impacto económico de los eventos adversos en los hospitales españoles a partir del Conjunto Mínimo Básico de Datos. *Gac Sanit*. febrero de 2014;28(1):48-54.
10. Jm A, de de M. Estudio Nacional de Efectos Adversos ligados a la Hospitalización (ENEAS).
11. Estudio APEAS. Estudio sobre la seguridad de los pacientes en atención primaria de salud. Madrid: Ministerio de Sanidad y Consumo; 2008.
12. Haerdlein A, Debold E, Rottenkolber M, Boehmer AM, Pudritz YM, Shahid F, et al. Which Adverse Events and Which Drugs Are Implicated in Drug-Related Hospital Admissions? A Systematic Review and Meta-Analysis. *J Clin Med*. 7 de febrero de 2023;12(4):1320.
13. Thomsen LA, Winterstein AG, SØndergaard B, HaugbØlle LS, Melander A. Systematic Review of the Incidence and Characteristics of Preventable Adverse Drug Events in Ambulatory Care. *Ann Pharmacother*. septiembre de 2007;41(9):1411-26.
14. Temporal Trends in Rates of Patient Harm from Medical Care. *N Engl J Med*. 17 de febrero de 2011;364(7):684-5.
15. Wolfe D, Yazdi F, Kanji S, Burry L, Beck A, Butler C, et al. Incidence, causes, and consequences of preventable adverse drug reactions occurring in inpatients: A systematic review of systematic reviews. Todd A, editor. *PLOS ONE*. 11 de octubre de 2018;13(10):e0205426.
16. Formica D, Sultana J, Cutroneo P, Lucchesi S, Angelica R, Crisafulli S, et al. The economic burden of preventable adverse drug reactions: a systematic review of observational studies. *Expert Opin Drug Saf*. 3 de julio de 2018;17(7):681-95.
17. Carrasco-Garrido P, de Andrés LA, Barrera VH, de Miguel GÁ, Jiménez-García R. Trends of adverse drug reactions related-hospitalizations in Spain (2001-2006). *BMC Health Serv Res*. 13 de octubre de 2010;10(1):287.

18. Agencia Española de medicamentos y productos sanitarios (AEMPS). SEFV-H Resumen de Actividad 2021. [Internet]. Disponible en: https://www.aemps.gob.es/vigilancia/medicamentosUsoHumano/SEFV-H/docs/resumen_actividad_SEFVH-2022.pdf
19. European Medicine Agency (EMA). EudraVigilance [Internet]. European Medicines Agency. 2018 [citado 13 de marzo de 2023]. Disponible en: <https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance>
20. Notificación electrónica de sospechas de Reacciones Adversas a Medicamentos por los profesionales sanitarios [Internet]. Agencia Española de Medicamentos y Productos Sanitarios. [citado 13 de abril de 2023]. Disponible en: https://www.aemps.gob.es/informa/notasinformativas/medicamentosusohumano-3/seguridad-1/2013/ni-muh_fv_02-2013-notifica-ram/
21. Ministerio de Sanidad y Consumo. Real Decreto 1344/2007, de 11 de octubre, por el que se regula la farmacovigilancia de medicamentos de uso humano [Internet]. Sec. 1, Real Decreto 1344/2007 nov 1, 2007 p. 44631-40. Disponible en: <https://www.boe.es/eli/es/rd/2007/10/11/1344>
22. Consejería de Sanidad DG de Gestión Económico-Financiera y Farmacia y DG de Inspección y Ordenación Sanitaria. Uso seguro de Medicamentos y Productos Sanitarios [Internet]. [citado 13 de marzo de 2023]. Disponible en: <https://seguridadmedicamento.sanidadmadrid.org/>
23. Karch FE, Lasagna L. Toward the operational identification of adverse drug reactions. Clin Pharmacol Ther. marzo de 1977;21(3):247-54.
24. Ekman E, Petersson G, Tågerud S, Bäckström M. Awareness among nurses about reporting of adverse drug reactions in Sweden. Drug Healthc Patient Saf. 28 de junio de 2012;4:61-6.
25. Granas AG, Buajordet M, Stenberg-Nilsen H, Harg P, Horn AM. Pharmacists' attitudes towards the reporting of suspected adverse drug reactions in Norway: PHARMACISTS AND ADVERSE DRUG REACTION REPORTING. Pharmacoepidemiol Drug Saf. abril de 2007;16(4):429-34.
26. Sandberg A, Salminen V, Heinonen S, Sivén M. Under-Reporting of Adverse Drug Reactions in Finland and Healthcare Professionals' Perspectives on How to Improve Reporting. Healthcare. 31 de mayo de 2022;10(6):1015.

27. Li R, Curtis K, Zaidi ST, Van C, Castelino R. A new paradigm in adverse drug reaction reporting: consolidating the evidence for an intervention to improve reporting. *Expert Opin Drug Saf.* septiembre de 2022;21(9):1193-204.
28. Fossouo Tagne J, Yakob RA, Mcdonald R, Wickramasinghe N. Linking Activity Theory Within User-Centered Design: Novel Framework to Inform Design and Evaluation of Adverse Drug Reaction Reporting Systems in Pharmacy. *JMIR Hum Factors.* 24 de febrero de 2023;10:e43529.
29. Forster AJ, Jennings A, Chow C, Leeder C, van Walraven C. A systematic review to evaluate the accuracy of electronic adverse drug event detection. *J Am Med Inform Assoc.* enero de 2012;19(1):31-8.
30. Spasic I, Nenadic G. Clinical Text Data in Machine Learning: Systematic Review. *JMIR Med Inform.* 31 de marzo de 2020;8(3):e17984.
31. Murphy RM, Klopotoska JE, de Keizer NF, Jager KJ, Leopold JH, Dongelmans DA, et al. Adverse drug event detection using natural language processing: A scoping review of supervised learning methods. Qamar U, editor. *PLOS ONE.* 3 de enero de 2023;18(1):e0279842.
32. Quazi S. Artificial intelligence and machine learning in precision and genomic medicine. *Med Oncol.* 15 de junio de 2022;39(8):120.
33. Anaya-Isaza A, Mera-Jiménez L, Zequera-Díaz M. An overview of deep learning in medical imaging. *Inform Med Unlocked.* 2021;26:100723.
34. Madrid C de. Éxito en la implantación del Módulo Único de Prescripción de la Comunidad de Madrid [Internet]. Comunidad de Madrid. 2019 [citado 13 de abril de 2023]. Disponible en: <https://www.comunidad.madrid/noticias/2019/10/24/exito-implantacion-modulo-unico-prescripcion-comunidad-madrid>
35. InfoREC 20. Receta Electrónica del Servicio Canario de la Salud [Internet]. [citado 13 de abril de 2023]. Disponible en: <https://www3.gobiernodecanarias.org/sanidad/scs/contenidoGenerico.jsp?idDocument=391b6238-e760-11e4-aa6f-cdf8d9a72f8&idCarpeta=1c178875-0973-11e3-8b8d-350345c703a5>
36. Servicio Gallego de Salud. Sistema inteligente de alertas multinivel - Consellería de Sanidade - Servizo Galego de Saúde [Internet]. [citado 13 de marzo de 2023]. Disponible en: <https://www.sergas.es/Hospital-2050---Innova-Saude/Sistema-intelixente-de-alertas-multinivel?idioma=es>

37. Premiada una herramienta informática del SNS-O que mejora la seguridad de la farmacoterapia [Internet]. [citado 11 de marzo de 2024]. Disponible en: https://www.navarra.es/home_es/Actualidad/Sala+de+prensa/Noticias/2018/01/02/Premiada+herramienta+informatica+del+SNS-O+que+mejora+seguridad+farmacoterapia.htm
38. Seguretat clínica a la recepta electrònica [Internet]. Medicaments i farmàcia. [citado 11 de marzo de 2024]. Disponible en: <http://medicaments.gencat.cat/ca/professionals/seguretat/seguretat-clinica-recepta-electronica/>
39. Oronoz M, Gojenola K, Pérez A, De Ilarraza AD, Casillas A. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. *J Biomed Inform.* agosto de 2015;56:318-32.
40. Santiso S, Pérez A, Casillas A, Oronoz M. Neural negated entity recognition in Spanish electronic health records. *J Biomed Inform.* mayo de 2020;105:103419.
41. Santiso S, Pérez A, Casillas A. Smoothing dense spaces for improved relation extraction between drugs and adverse reactions. *Int J Med Inf.* agosto de 2019;128:39-45.
42. Higgins JPT, Green S. Manual Cochrane de revisiones sistemáticas de intervenciones. Cochrane. 2011;(March):1-639.
43. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ.* 2021;372.
44. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ Online.* 2017;358:j4008-j4008.
45. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ.* 28 de agosto de 2019;14898.
46. JBI Critical Appraisal Tools | JBI [Internet]. [citado 24 de noviembre de 2023]. Disponible en: <https://jbi.global/critical-appraisal-tools>
47. Zotero. Zotero: Your personal research assistant. 2023.
48. Borissov N, Haas Q, Minder B, Kopp-Heim D, von Gernler M, Janka H, et al. Reducing systematic review burden using Deduklick: a novel, automated, reliable, and explainable deduplication algorithm to foster medical research. *Syst Rev.* 17 de agosto de 2022;11(1):172.

49. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programme. Third edition. Oxford: Oxford University Press; 2005.
50. López de Argumedo, M, Reviriego, E, Gutiérrez, A, Bayón, J. Actualización del Sistema de Trabajo Compartido para Revisiones Sistemáticas de la Evidencia Científica y Lectura Crítica (Plataforma FLC 3.0). OSTEBA; 2017. (Informes de Evaluación de Tecnologías Sanitarias).
51. O'Rourke B, Oortwijn W, Schuller T, International Joint Task Group. The new definition of health technology assessment: A milestone in international collaboration. *Int J Technol Assess Health Care*. junio de 2020;36(3):187-90.
52. Peters MDJ, Godfrey C, McInerney P, Khalil H, Larsen P, Marnie C, et al. Best practice guidance and reporting items for the development of scoping review protocols. *JBIM Evid Synth*. 1 de abril de 2022;20(4):953-68.
53. EUnetHTA Joint Action 2 Work Package 8. HTA Core Model ® Version 3.0 [Internet]. 2016. Disponible en: www.htacoremodel.info/BrowseModel.aspx
54. Moher J, Oxman AD, Rosenbaum S, Schünemann HJ, Guyatt G, Flottorp S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Health Res Policy Syst*. 29 de mayo de 2018;16(1):45.
55. Van Der Wilt GJ, Bloemen B, Grin J, Gutierrez-Ibarluzea I, Sampietro-Colom L, Refolo P, et al. Integrating Empirical Analysis and Normative Inquiry in Health Technology Assessment: The Values in Doing Assessments of Health Technologies Approach. *Int J Technol Assess Health Care*. 23 de junio de 2022;38(1):e52.
56. Gonzalez-Gil T. Plantilla para ayudarte a entender un estudio cualitativo. En: CASPe. Guías CASPe de Lectura Crítica de la Literatura Médica. Alicante: CASPe; 2010. Cuaderno III. p.3-8. 2019.
57. Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev*. 2019;4(1):1-7.
58. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Educ Inf*. 2018;34(4):285-91.
59. Bagattini F, Karlsson I, Rebane J, Papapetrou P. A classification framework for exploiting sparse multi-variate temporal features

- with application to adverse drug event detection in medical records. *BMC Med Inform Decis Mak.* diciembre de 2019;19(1):7.
60. Banerji A, Lai KH, Li Y, Saff RR, Camargo CA, Blumenthal KG, et al. Natural Language Processing Combined with ICD-9-CM Codes as a Novel Method to Study the Epidemiology of Allergic Drug Reactions. *J Allergy Clin Immunol Pract.* marzo de 2020;8(3):1032-1038.e1.
 61. Chan S, Tham M, Tan S, Loke C, Foo B, Fan Y, et al. Development and Validation of Algorithms for the Detection of Statin Myopathy Signals From Electronic Medical Records. *Clin Pharmacol Ther.* mayo de 2017;101(5):667-74.
 62. Chen C, Jia W, Guo D, Zhu M, Xu Y, Wang X, et al. Development of a Computer-Assisted Adverse Drug Events Alarm and Assessment System for Hospital Inpatients in China. *Ther Innov Regul Sci.* enero de 2020;54(1):32-41.
 63. Iivanainen S, Ekstrom J, Virtanen H, Kataja VV, Koivunen JP. Electronic patient-reported outcomes and machine learning in predicting immune-related adverse events of immune checkpoint inhibitor therapies. *BMC Med Inform Decis Mak.* diciembre de 2021;21(1):205.
 64. Inglis JM, Bacchi S, Troelnikov A, Smith W, Shakib S. Automation of penicillin adverse drug reaction categorisation and risk stratification with machine learning natural language processing. *Int J Med Inf.* diciembre de 2021;156:104611.
 65. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register. Ouzounis CA, editor. *PLOS ONE.* 14 de agosto de 2015;10(8):e0134208.
 66. Iqbal E, Mallah R, Rhodes D, Wu H, Romero A, Chang N, et al. ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records. Groza T, editor. *PLOS ONE.* 9 de noviembre de 2017;12(11):e0187121.
 67. Jeon E, Kim Y, Park H, Park RW, Shin H, Park HA. Analysis of Adverse Drug Reactions Identified in Nursing Notes Using Reinforcement Learning. *Healthc Inform Res.* 30 de abril de 2020;26(2):104-11.
 68. McMaster C, Chan J, Liew DFL, Su E, Frauman AG, Chapman WW, et al. Developing a deep learning natural language processing algorithm for automated reporting of adverse drug reactions. *J Biomed Inform.* enero de 2023;137:104265.

69. McMaster C, Liew D, Keith C, Aminian P, Frauman A. A Machine-Learning Algorithm to Optimise Automated Adverse Drug Reaction Detection from Clinical Coding. *Drug Saf.* junio de 2019;42(6):721-5.
70. Pandey C, Ibrahim Z, Wu H, Iqbal E, Dobson R. Improving RNN with Attention and Embedding for Adverse Drug Reactions. En: *Proceedings of the 2017 International Conference on Digital Health* [Internet]. London United Kingdom: ACM; 2017 [citado 21 de diciembre de 2023]. p. 67-71. Disponible en: <https://dl.acm.org/doi/10.1145/3079452.3079501>
71. Quillet A, Colin O, Bourgeois N, Favrelière S, Ferru A, Boinot L, et al. Detection of adverse drug reactions: evaluation of an automatic data processing applied in oncology performed in the French Diagnosis Related Groups database. *Fundam Clin Pharmacol.* abril de 2018;32(2):227-33.
72. Santiso S, Pérez A, Casillas A. Adverse Drug Reaction extraction: Tolerance to entity recognition errors and sub-domain variants. *Comput Methods Programs Biomed.* febrero de 2021;199:105891.
73. Smith J, Chen Q, Denny J, Roden D, Johnson K, Miller R. Evaluation of a Novel System to Enhance Clinicians' Recognition of Preadmission Adverse Drug Reactions. *Appl Clin Inform.* abril de 2018;09(02):313-25.
74. Tang H, Solti I, Kirkendall E, Zhai H, Lingren T, Meller J, et al. Leveraging Food and Drug Administration Adverse Event Reports for the Automated Monitoring of Electronic Health Records in a Pediatric Hospital. *Biomed Inform Insights.* enero de 2017;9:117822261771301.
75. Tang Y, Yang J, Ang PS, Dorajoo SR, Foo B, Soh S, et al. Detecting adverse drug reactions in discharge summaries of electronic medical records using Readpeer. *Int J Med Inf.* agosto de 2019;128:62-70.
76. Van Laar SA, Gombert-Handoko KB, Guchelaar H, Zwaveling J. An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma. *Clin Pharmacol Ther.* septiembre de 2020;108(3):644-52.
77. Yang J, Wang L, Phadke NA, Wickner PG, Mancini CM, Blumenthal KG, et al. Development and Validation of a Deep Learning Model for Detection of Allergic Reactions Using Safety Event Reports Across Hospitals. *JAMA Netw Open.* 16 de noviembre de 2020;3(11):e2022836.

78. Zhao J, Henriksson A, Asker L, Boström H. Predictive modeling of structured electronic health records for adverse drug event detection. *BMC Med Inform Decis Mak.* diciembre de 2015;15(S4):S1.
79. Muzaffar AF, Abdul-Massih S, Stevenson JM, Alvarez-Arango S. Use of the Electronic Health Record for Monitoring Adverse Drug Reactions. *Curr Allergy Asthma Rep.* julio de 2023;23(7):417-26.
80. Chen R, Zhang Y, Dou Z, Chen F, Xie K, Wang S. Data Sharing and Privacy in Pharmaceutical Studies. *Curr Pharm Des.* 2021;27(7):911-8.
81. Amin A, Malik M, Hussain A. EVOLVING TRENDS AND FUTURE PROSPECTS OF ADR MONITORING AND REPORTING: A LITERATURE REVIEW. 1 de septiembre de 2016 [citado 24 de noviembre de 2023]; Disponible en: <https://www.growkudos.com/publications/10.22159%252Fajpcr.2016.v9i5.12725>
82. Edrees H, Song W, Syrowatka A, Simona A, Amato MG, Bates DW. Intelligent Telehealth in Pharmacovigilance: A Future Perspective. *Drug Saf.* 1 de mayo de 2022;45(5):449-58.
83. Li R, Curtis K, Van C, Tabish Razi Zaidi S, Yeo CY, Arun Kali C, et al. Why hospital-based healthcare professionals do not report adverse drug reactions: a mixed methods study using the Theoretical Domains Framework. *Eur J Clin Pharmacol.* julio de 2022;78(7):1165-75.
84. Tillman EM, Suppes SL, Feldman K, Goldman JL. Enhancing Pediatric Adverse Drug Reaction Documentation in the Electronic Medical Record. *J Clin Pharmacol.* febrero de 2021;61(2):181-6.
85. Hurdle JF, Weir CR, Roth B, Hoffman J, Nebeker JR. Critical Gaps in the World's Largest Electronic Medical Record: Ad Hoc Nursing Narratives and Invisible Adverse Drug Events. *AMIA Annu Symp Proc.* 2003;2003:309-17.
86. Glassman PA, Belperio P, Simon B, Lanto A, Lee M. Exposure to automated drug alerts over time: effects on clinicians' knowledge and perceptions. *Med Care.* marzo de 2006;44(3):250-6.
87. Digitalidades. Guía «Inteligencia artificial ética en sanidad» de Digitales – DIGITALIDADES [Internet]. 2022 [citado 24 de noviembre de 2023]. Disponible en: <https://www.digitalidades.org/guia-inteligencia-artificial-etica-en-sanidad-de-digitales/>
88. Modelo de autoevaluación PIO – OEIAC [Internet]. [citado 24 de noviembre de 2023]. Disponible en: <https://oeiac.cat/es/modelo-de-autoevaluacion-pio/>

89. AI Research Sesgos Covid-19 · UAB Team for AI Research [Internet]. 2022 [citado 24 de noviembre de 2023]. Disponible en: <https://airesearch-sesgos-covid19.com/>
90. Lokmic-Tomkins Z, Davies S, Block LJ, Cochrane L, Dorin A, Von Gerich H, et al. Assessing the carbon footprint of digital health interventions: a scoping review. *J Am Med Inform Assoc*. 14 de noviembre de 2022;29(12):2128-39.
91. García-Berná JA, Fernández-Alemán JL, Carrillo De Gea JM, Toval A, Mancebo J, Calero C, et al. Energy efficiency in software: A case study on sustainability in personal health records. *J Clean Prod*. febrero de 2021;282:124262.
92. COMET Initiative | Core Outcome Set for patient safety in perioperative care: SAFEST Project [Internet]. [citado 21 de diciembre de 2023]. Disponible en: <https://www.comet-initiative.org/Studies/Details/2131>
93. COMET Initiative | Development of a core outcome set for the evaluation of a computerized physician order entry (CPOE) system [Internet]. [citado 21 de diciembre de 2023]. Disponible en: <https://www.comet-initiative.org/Studies/Details/2613>
94. Irujo M, Beitia G, Bes-Rastrollo M, Figueiras A, Hernández-Díaz S, Lasheras B. Factors that influence under-reporting of suspected adverse drug reactions among community pharmacists in a Spanish region. *Drug Saf*. 2007;30(11):1073-82.
95. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *N Engl J Med*. 27 de agosto de 2020;383(9):874-82.
96. Buslón N, Cortés A, Catuara-Solarz S, Cirillo D, Rementeria MJ. Raising awareness of sex and gender bias in artificial intelligence and health. *Front Glob Womens Health*. 6 de septiembre de 2023;4:970312.

Anexos

Anexo 1. Estrategias de búsqueda

Efectividad, seguridad y coste-efectividad

MEDLINE		
1	Drug-Related Side Effects and Adverse Reactions/di, cl, ep.	7276
2	Medication Errors/pc	6975
3	((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) adj (incident* or event* or reaction*)).ti,ab,kf.	31677
4	1 or 2 or 3	42728
5	Adverse Drug Reaction Reporting Systems/	8796
6	((("Drug safety" or "safety signal*" or "medicine use" or "reporting system*" or adverse or medication or "drug-associated" or "drugs associated" or ADE or ADR or AERS) and (surveillanc* or prevent* or detec* or predict* or identification or discovery or automated or method* or computer* or reporting or monitor* or pharmacovigilanc*)).ti.	16576
7	5 or 6	22859
8	Drug Monitoring/mt or Computer Systems/ or "Diagnosis, Computer-Assisted"/ or Knowledge bases/	49613
9	4 and 7 and 8	182
10	Databases, Factual/ and (Area Under Curve/ or "Sensitivity and Specificity"/ or "Predictive value of tests"/ or Program evaluation/ or *Pharmacoepidemiology/ or *Information Systems/)	5976
11	4 or 7	59152
12	10 and 11	142
13	Electronic Health Records/ or "Medical Records "/ or "Medical Informatics"/ or "Hospital Information Systems"/ or "Information Storage and Retrieval"/ or ("electronic* record*" or "electronic* medical*" or "electronic* data" or "text data" or "basic data set" or "medical record*" or "emr" or "health* record*" or "ehr" or "management system*" or "information system*" or "real-world data*" or "administrative data*" or "hospital data*" or "patient history" or "patient report*" or "extracted information" or "data extraction" or "large data sets" or "data sources" or "unstructured data" or "available data" or "coded data" or "Computer Systems" or "Automated" or "Information Storage" or "Information Storage retrieval").ab.	648378
14	Supervised Machine Learning/ or "Machine Learning"/ or "Natural Language Processing"/ or "Algorithms"/ or "Data Mining"/ or "Deep Learning"/ or Automation/ or "Pattern Recognition, Automated"/ or ("intelligen*" or "software*" or "health technology" or "ai based" or "Pattern Recognition" or "nlp" or "natural language" or "text-based" or "text-mining" or "clinical-text" or "minimum basic data set" or "mining*" or "machine*" or "learning methods" or "deep learning" or "artificial" or "mobile*" or "algorithm" or "neural network*" or "pharmacovigilance method*" or "multiple sources").ab.	1323884
15	4 and 7 and 13 and 14	347
16	9 or 12 or 15	645

Embase		
#1	'adverse drug reaction'/mj	125261
#2	medication error'/dm_pc	2500
#3	((('adverse drug' OR 'drug adverse' OR 'drug related' OR ade OR adr OR ades OR adrs OR 'intraoperative adverse' OR 'immune-related adverse' OR 'adr related hospitalization' OR medication) NEAR/1 (incident* OR event* OR reaction*)):ti,ab,kw	53487
#4	#1 OR #2 OR #3	57467
#5	'pharmacovigilance'/mj	2406
#6	('drug safety':ti OR 'safety signal':ti OR 'medicine use':ti OR 'reporting system':ti OR adverse:ti OR medication:ti OR 'drug-associated':ti OR 'drugs associated':ti OR ade:ti OR adr:ti OR aers:ti) AND (surveillanc*:ti OR prevent*:ti OR detec*:ti OR predict*:ti OR identification:ti OR discovery:ti OR automated:ti OR method*:ti OR computer*:ti OR reporting:ti OR monitor*:ti OR pharmacovigilanc*:ti)	25196
#7	#5 OR #6	29035
#8	'drug monitoring'/mj OR 'computer system'/mj OR 'computer assisted diagnosis'/mj OR 'knowledge base'/mj	54162
#9	#4 AND #7 AND #8	345
#10	'factual database'/exp AND ('area under the curve'/exp OR 'sensitivity and specificity'/exp OR 'predictive value'/exp OR 'program evaluation'/exp OR 'pharmacoepidemiology'/exp OR 'information system'/exp)	7254
#11	#4 OR #7	80052
#12	#10 AND #11	69
#13	'electronic health record'/exp OR 'medical record'/exp OR 'medical informatics'/exp OR 'hospital information system'/exp OR 'information retrieval'/exp OR 'electronic* record':ab OR 'electronic* medical':ab OR 'electronic* data':ab OR 'text data':ab OR 'basic data set':ab OR 'medical record':ab OR 'emr':ab OR 'health* record':ab OR 'ehr':ab OR 'management system':ab OR 'information system':ab OR 'real-world data':ab OR 'administrative data':ab OR 'hospital data':ab OR 'patient history':ab OR 'patient report':ab OR 'extracted information':ab OR 'data extraction':ab OR 'large data sets':ab OR 'data sources':ab OR 'unstructured data':ab OR 'available data':ab OR 'coded data':ab OR 'computer systems':ab OR 'automated':ab OR 'information storage':ab OR 'information storage retrieval':ab	1042428
#14	'supervised machine learning'/exp OR 'machine learning'/exp OR 'natural language processing'/exp OR 'algorithm'/exp OR 'data mining'/exp OR 'deep learning'/exp OR 'automation'/exp OR 'automated pattern recognition'/exp OR 'intelligen':ab OR 'software':ab OR 'health technology':ab OR 'ai based':ab OR 'pattern recognition':ab OR 'nlp':ab OR 'natural language':ab OR 'text-based':ab OR 'text-mining':ab OR 'clinical-text':ab OR 'minimum basic data set':ab OR 'mining':ab OR 'machine':ab OR 'learning methods':ab OR 'deep learning':ab OR 'artificial':ab OR 'mobile':ab OR 'algorithm':ab OR 'neural network':ab OR 'pharmacovigilance method':ab OR 'multiple sources':ab	1942877
#15	#4 AND #7 AND #13 AND #14	345
#16	#9 OR #12 OR #15	682
#17	#16 AND [embase]/lim NOT ([embase]/lim AND [medline]/lim)	214

CENTRAL / CDSR		
#1	MeSH descriptor: [Drug-Related Side Effects and Adverse Reactions] this term only and with qualifier(s): [classification - CL, diagnosis - DI, epidemiology - EP]	735
#2	MeSH descriptor: [Medication Errors] this term only and with qualifier(s): [prevention & control - PC]	234

CENTRAL / CDSR		
#3	((('adverse drug' OR 'drug adverse' OR 'drug related' OR ade OR adr OR ades OR adrs OR 'intraoperative adverse' OR 'immune-related adverse' OR 'adr related hospitalization' OR medication) NEAR/1 (incident* OR event* OR reaction*)):ti,ab,kw	196057
#4	#1 OR #2 OR #3	196235
#5	MeSH descriptor: [Adverse Drug Reaction Reporting Systems] this term only	157
#6	('drug safety':ti OR 'safety signal':ti OR 'medicine use':ti OR 'reporting system*':ti OR adverse:ti OR medication:ti OR 'drug-associated':ti OR 'drugs associated':ti OR ade:ti OR adr:ti OR aers:ti) AND (surveillanc*:ti OR prevent*:ti OR detec*:ti OR predict*:ti OR identification:ti OR discovery:ti OR automated:ti OR method*:ti OR computer*:ti OR reporting:ti OR monitor*:ti OR pharmacovigilanc*:ti)	4016
#7	#5 OR #6	4125
#8	[mh ^"Drug Monitoring"/mt] OR [mh ^"Computer System"] OR [mh ^"Computer assisted, diagnosis"] OR [mh ^"Knowledge base"]	1136
#9	#4 AND #7 AND #8	17
#10	[mh ^"Databases, Factual"] and ([mh ^"Area Under Curve"] or [mh ^"Sensitivity and Specificity"] or [mh ^"Predictive value of tests"] or [mh ^"Program evaluation"] or [mh ^"Pharmacoepidemiology"] or [mh ^"Information Systems"])	68
#11	#4 OR #7	198838
#12	#10 AND #11	0
#13	[mh ^"Electronic Health Records"] or [mh ^"Medical Records"] or [mh ^"Medical Informatics"] or [mh ^"Hospital Information Systems"] or [mh ^"Information Storage and Retrieval"] or ("electronic* record*" or "electronic* medical*" or "electronic* data" or "text data" or "basic data set" or "medical record*" or "emr" or "health* record*" or "ehr" or "management system*" or "information system*" or "real-world data*" or "administrative data*" or "hospital data*" or "patient history" or "patient report*" or "extracted information" or "data extraction" or "large data sets" or "data sources" or "unstructured data" or "available data" or "coded data" or "Computer Systems" or "Automated" or "Information Storage" or "Information Storage retrieval"):ab	28198
#14	[mh ^"Supervised Machine Learning"] or [mh ^"Machine Learning"] or [mh ^"Natural Language Processing"] or [mh ^"Algorithms"] or [mh ^"Data Mining"] or [mh ^"Deep Learning"] or [mh ^"Automation"] or [mh ^"Pattern Recognition, Automated"] or ("intelligen*" or "software*" or "health technology" or "ai based" or "Pattern Recognition" or "nlp" or "natural language" or "text-based" or "text-mining" or "clinical-text" or "minimum basic data set" or "mining*" or "machine*" or "learning methods" or "deep learning" or "artificial" or "mobile*" or "algorithm*" or "neural network*" or "pharmacovigilance method*" or "multiple sources"):ab	61444
#15	#4 AND #7 AND #13 AND #14	14
#16	#9 OR #12 OR #15 en Revisiones Cochrane, Protocolos Cochrane	3
#17	#9 OR #12 OR #15 en Ensayos	28
#18	#9 OR #12 OR #15	31

Web of Science Core Collection		
#1	((TI=((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) NEAR/1 (incident* or event* or reaction*)))) OR AB=((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) NEAR/1 (incident* or event* or reaction*)))) OR AK=((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) NEAR/1 (incident* or event* or reaction*)))) OR KP=((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or	42612

Web of Science Core Collection		
	adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) NEAR/1 (incident* or event* or reaction*))	
#2	TI=((("Drug safety" or "safety signal" or "medicine use" or "reporting system*" or adverse or medication or "drug-associated" or "drugs associated" or ADE or ADR or AERS) and (surveillanc* or prevent* or detec* or predict* or identification or discovery or automated or method* or computer* or reporting or monitor* or pharmacovigilanc*))AERS) and (surveillanc* or prevent* or detec* or predict* or identification or discovery or automated or method* or computer* or reporting or monitor* or pharmacovigilanc*))	29785
#3	AB=((("electronic* record*" or "electronic* medical*" or "electronic* data" or "text data" or "basic data set" or "medical record*" or "emr" or "health* record*" or "ehr" or "management system*" or "information system*" or "real-world data*" or "administrative data*" or "hospital data*" or "patient history" or "patient report*" or "extracted information" or "data extraction" or "large data sets" or "data sources" or "unstructured data" or "available data" or "coded data" or "Computer Systems" or "Automated" or "Information Storage" or "Information Storage retrieval"))	972943
#4	AB=((("intelligen*" or "software*" or "health technology" or "ai based" or "Pattern Recognition" or "nlp" or "natural language" or "text-based" or "text-mining" or "clinical-text" or "minimum basic data set" or "learning methods" or "deep learning" or "artificial" or "neural network*" or "pharmacovigilance method*" or "multiple sources"))	2088377
#5	#1 OR #2	66783
#6	#3 AND #4	136116
#7	#5 AND #6	470

Aspectos éticos, legales, organizativos, sociales y ambientales

MEDLINE		
1	Drug-Related Side Effects and Adverse Reactions/di, cl, ep	7287
2	Medication Errors/pc	6981
3	((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) adj (incident* or event* or reaction*)).ti,ab,kf.	31853
4	1 or 2 or 3	42912
5	Adverse Drug Reaction Reporting Systems/	8804
6	((("Drug safety" or "patient safety" or "safety signal" or "medicine use" or "reporting system" or adverse or medication or "drug-associated" or "drugs associated" or ADE or ADR or AERS) and (surveillanc* or prevent* or detec* or predict* or identification or discovery or automated or method* or computer* or reporting or monitor* or pharmacovigilanc* or alert*)).ti.	17705
7	5 or 6	23977
8	Drug Monitoring/mt or Computer Systems/ or "Diagnosis, Computer-Assisted"/ or Knowledge bases/ or "Decision Support Systems, Clinical"/	58275
9	4 and 7 and 8	342
10	Databases, Factual/ and (Area Under Curve/ or "Sensitivity and Specificity"/ or "Predictive value of tests"/ or Program evaluation/ or "Pharmacoepidemiology/ or "Information Systems/)	5974
11	4 or 7	60311
12	10 and 11	144
13	Electronic Health Records/ or "Medical Records "/ or "Medical Informatics"/ or "Hospital Information Systems"/ or "Information Storage and Retrieval"/ or ("electronic* record*" or "electronic* medical*" or "electronic* data" or "text data" or "basic data set" or "medical record*" or "emr" or "health* record*" or "ehr" or "management system*" or "information system*" or "real-world data*" or "administrative data*" or "hospital data*" or "patient history" or "patient report*" or "extracted information" or "data extraction" or "large data sets" or "data sources" or "unstructured data" or "available data" or "coded data" or "Computer Systems" or "Automated" or "Information Storage" or "Information Storage retrieval").ab.	650419
14	Supervised Machine Learning/ or "Machine Learning"/ or "Natural Language Processing"/ or "Algorithms"/ or "Data Mining"/ or "Deep Learning"/ or Automation/ or "Pattern Recognition, Automated"/ or ("intelligen*" or "software*" or "health technology" or "ai based" or "Pattern Recognition" or "nlp" or "natural language" or "text-based" or "text-mining" or "clinical-text" or "minimum basic data set" or "mining*" or "machine*" or "learning methods" or "deep learning" or "artificial" or "mobile*" or "algorithm*" or "neural network*" or "pharmacovigilance method*" or "multiple sources").ab.	1328973
15	11 and 13 and 14	1209
16	9 or 12 or 17	1645
17	((("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or structured or guide) adj3 (interview* or discussion* or questionnaire*)) or (focus group* or qualitative or ethnograph* or fieldwork or "field work" or "key informant" or metasynthes* or meta-synthes* or metasummar* or meta-summar* or metastud* or meta-stud* or metathem* or meta-them* or meta-ethno* or emic or etic or phenomenolog* or grounded theory or constant compar* or (thematic* adj3 analys*) or theoretical sampl* or purposive sampl* or hermeneutic*)).ti,ab. or Interviews as Topic/ or focus groups/ or narration/ or Qualitative Research/ or Grounded Theory/	551377
18	(user* experienc* or User-Centered or user-focused or ((user* or professional* or physician*) adj2 (view* or satisfaction or feedback or interface or experienc*))) or privacy or cybersecurity or feasibility or viability or practicability or acceptability or acceptance or approval or organizational or facilitator* or motivator* or barrier* or mixed-method*).ti,ab.	1172407
19	(equity or equality or fairness or ethical issues or ethical dilemmas or ethical problems or ethical challenges or ethic* or bias or prejudice or partiality).ti,ab.	418970

MEDLINE		
20	(fatigue adj3 alert*).ti,ab.	483
21	19 or 20 or 21 or 22	1955116
22	18 and 23	320

Embase		
#1	'adverse drug reaction'/mj	125321
#2	'medication error'/dm_pc	2500
#3	((('adverse drug' OR 'drug adverse' OR 'drug related' OR ade OR adr OR ades OR adrs OR 'intraoperative adverse' OR 'immune-related adverse' OR 'adr related hospitalization' OR medication) NEAR/1 (incident* OR event* OR reaction*)):ti,ab,kw	53877
#4	#1 OR #2 OR #3	175236
#5	'pharmacovigilance'/mj	2470
#6	('drug safety':ti OR 'drug safety':ti OR 'safety signal':ti OR 'medicine use':ti OR 'reporting system':ti OR adverse:ti OR medication:ti OR 'drug-associated':ti OR 'drugs associated':ti OR ade:ti OR adr:ti OR aers:ti) AND (surveillanc*:ti OR prevent*:ti OR detec*:ti OR predict*:ti OR identification:ti OR discovery:ti OR automated:ti OR method*:ti OR computer*:ti OR reporting:ti OR monitor*:ti OR pharmacovigilanc*:ti OR alert:ti)	25486
#7	#5 OR #6	27199
#8	'drug monitoring'/mj OR 'computer system'/mj OR 'computer assisted diagnosis'/mj OR 'knowledge base'/mj OR 'clinical decision support system'/mj	56987
#9	#4 AND #7 AND #8	300
#10	'factual database'/exp AND ('area under the curve'/exp OR 'sensitivity and specificity'/exp OR 'predictive value'/exp OR 'program evaluation'/exp OR 'pharmacoepidemiology'/exp OR 'information system'/exp)	7315
#11	#4 OR #7	195748
#12	#10 AND #11	71
#13	'electronic health record'/exp OR 'medical record'/exp OR 'medical informatics'/exp OR 'hospital information system'/exp OR 'information retrieval'/exp OR 'electronic record':ab OR 'electronic* medical':ab OR 'electronic* data':ab OR 'text data':ab OR 'basic data set':ab OR 'medical record':ab OR 'emr':ab OR 'health* record':ab OR 'ehr':ab OR 'management system':ab OR 'information system':ab OR 'real-world data':ab OR 'administrative data':ab OR 'hospital data':ab OR 'patient history':ab OR 'patient report':ab OR 'extracted information':ab OR 'data extraction':ab OR 'large data sets':ab OR 'data sources':ab OR 'unstructured data':ab OR 'available data':ab OR 'coded data':ab OR 'computer systems':ab OR 'automated':ab OR 'information storage':ab OR 'information storage retrieval':ab	1042428
#14	'supervised machine learning'/exp OR 'machine learning'/exp OR 'natural language processing'/exp OR 'algorithm'/exp OR 'data mining'/exp OR 'deep learning'/exp OR 'automation'/exp OR 'automated pattern recognition'/exp OR 'intelligen':ab OR 'software':ab OR 'health technology':ab OR 'ai based':ab OR 'pattern recognition':ab OR 'nlp':ab OR 'natural language':ab OR 'text-based':ab OR 'text-mining':ab OR 'clinical-text':ab OR 'minimum basic data set':ab OR 'mining':ab OR 'machine':ab OR 'learning methods':ab OR 'deep learning':ab OR 'artificial':ab OR 'mobile':ab OR 'algorithm':ab OR 'neural network':ab OR 'pharmacovigilance method':ab OR 'multiple sources':ab	1957062
#15	#4 AND #7 AND #13 AND #14	353
#16	#9 OR #15	642
#17	#11 AND #13 AND #14	1856
#18	#9 OR #12 OR #17	22060
#19	((('semi-structured' OR semistructured OR unstructured OR informal OR 'in-depth' OR indepth OR 'face-to-face' OR structured OR guide) NEAR/3 (interview* OR discussion* OR questionnaire)):ti,ab) OR 'focus group':ti,ab OR qualitative:ti,ab OR	602084

Embase		
	ethnograph*:ti,ab OR fieldwork:ti,ab OR 'field work':ti,ab OR 'key informant':ti,ab OR metasynthes*:ti,ab OR 'meta syntheses':ti,ab OR metasummar*:ti,ab OR 'meta summar':ti,ab OR metastud*:ti,ab OR 'meta stud':ti,ab OR metathem*:ti,ab OR 'meta them':ti,ab OR 'meta ethno':ti,ab OR emic:ti,ab OR etic:ti,ab OR phenomenolog*:ti,ab OR 'grounded theory':ti,ab OR 'constant compar*:ti,ab OR ((thematic* NEAR/3 analys'):ti,ab) OR 'theoretical sampl':ti,ab OR 'purposive sampl':ti,ab OR hermeneutic*:ti,ab OR 'interview'/mj OR 'focus group'/mj OR 'verbal communication'/mj OR 'qualitative research'/mj OR 'grounded theory'/mj	
#20	'user* experienc*:ti,ab OR 'user centered':ti,ab OR 'user focused':ti,ab OR ((user* OR professional* OR physician*) NEAR/2 (view* OR satisfaction OR feedback OR interface OR experienc*)):ti,ab) OR privacy:ti,ab OR cybersecurity:ti,ab OR feasibility:ti,ab OR viability:ti,ab OR practicability:ti,ab OR acceptability:ti,ab OR acceptance:ti,ab OR approval:ti,ab OR organizational:ti,ab OR facilitator*:ti,ab OR motivator*:ti,ab OR barrier*:ti,ab OR 'mixed method*:ti,ab	1548630
#21	equity:ti,ab OR equality:ti,ab OR fairness:ti,ab OR dilemmas:ti,ab OR challenges:ti,ab OR ethic*:ti,ab OR bias:ti,ab OR prejudice:ti,ab OR partiality:ti,ab	983586
#22	(fatigue NEAR/3 alert*):ti,ab	737
#23	#19 OR #20 OR #21 OR #22	2727633
#24	#18 AND #23	449

Web of Science		
1	((TI=((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) NEAR/1 (incident* or event* or reaction*)))) OR AB=((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) NEAR/1 (incident* or event* or reaction*)))) OR AK=((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) NEAR/1 (incident* or event* or reaction*)))) OR KP=((("adverse drug" or "drug adverse" or "drug related" or ade or adr or ades or adrs or "intraoperative adverse" or "immune-related adverse" or "adr related hospitalization" or medication) NEAR/1 (incident* or event* or reaction*)))	42856
2	TI=((("Drug safety" or "Patient safety" or "safety signal" or "medicine use" or "reporting system" or adverse or medication or "drug-associated" or "drugs associated" or ADE or ADR or AERS) and (surveillanc* or prevent* or detec* or predict* or identification or discovery or automated or method* or computer* or reporting or monitor* or pharmacovigilanc* OR Alert)))	30167
3	AB=((("electronic* record" or "electronic* medical" or "electronic* data" or "text data" or "basic data set" or "medical record" or "emr" or "health* record" or "ehr" or "management system" or "information system" or "real-world data" or "administrative data" or "hospital data" or "patient history" or "patient report" or "extracted information" or "data extraction" or "large data sets" or "data sources" or "unstructured data" or "available data" or "coded data" or "Computer Systems" or "Automated" or "Information Storage" or "Information Storage retrieval"))	978157
4	AB=((("intelligen" or "software" or "health technology" or "ai based" or "Pattern Recognition" or "nlp" or "natural language" or "text-based" or "text-mining" or "clinical-text" or "minimum basic data set" or "learning methods" or "deep learning" or "artificial" or "neural network" or "pharmacovigilance method" or "multiple sources" OR CDSS))	2105071
5	#1 OR #2	67317
6	#3 AND #4	137474
7	#6 AND #5	504
8	TI=((("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or structured or guide) NEAR/3 (interview* or discussion* or questionnaire*)) or (focus group* or qualitative or ethnograph* or fieldwork or "field	1849537

Web of Science

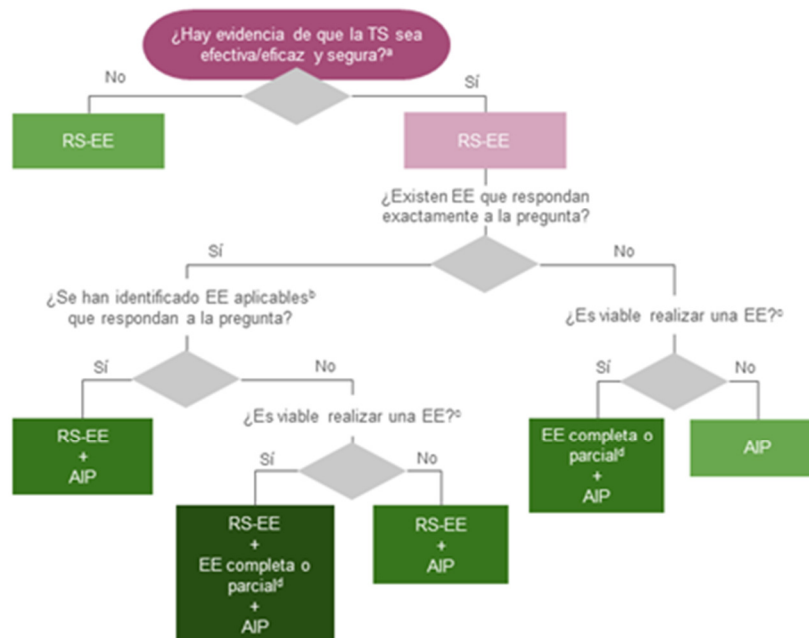
	work" or "key informant" or metasynthes* or meta-synthes* or metasummar* or meta-summar* or metastud* or meta-stud* or metathem* or meta-them* or meta-ethno* or emic or etic or phenomenolog* or grounded theory or constant compar* or (thematic* NEAR/3 analys*) or "theoretical sampl*" or "purposive sampl*" or hermeneutic*)) OR AB=((("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or structured or guide) NEAR/3 (interview* or discussion* or questionnaire*)) or (focus group* or qualitative or ethnograph* or fieldwork or "field work" or "key informant" or metasynthes* or meta-synthes* or metasummar* or meta-summar* or metastud* or meta-stud* or metathem* or meta-them* or meta-ethno* or emic or etic or phenomenolog* or grounded theory or constant compar* or (thematic* NEAR/3 analys*) or "theoretical sampl*" or "purposive sampl*" or hermeneutic*))	
9	TI=(user* experienc* or User-Centered or user-focused or ((user* or professional* or physician*) NEAR/2 (view* or satisfaction or feedback or interface or experienc*)) or privacy or cybersecurity or feasibility or viability or practicability or acceptability or acceptance or approval or organizational or facilitator* or motivator* or barrier* or mixed-method*) OR AB=(user* experienc* or User-Centered or user-focused or ((user* or professional* or physician*) NEAR/2 (view* or satisfaction or feedback or interface or experienc*)) or privacy or cybersecurity or feasibility or viability or practicability or acceptability or acceptance or approval or organizational or facilitator* or motivator* or barrier* or mixed-method*)	2508477
10	TI=(equity or equality or fairness or ethical issues or ethical dilemmas or ethical problems or ethical challenges or ethic* or bias or prejudice or partiality) OR AB=(equity or equality or fairness or ethical issues or ethical dilemmas or ethical problems or ethical challenges or ethic* or bias or prejudice or partiality)	1162487
11	TI=(fatigue NEAR/3 alert*) OR AB=(fatigue NEAR/3 alert*)	672
12	#11 OR #10 OR #9 OR #8	5129892
13	#12 AND #7	128

Anexo 2. Algoritmo para la realización de evaluación económica en Evaluación de Tecnologías Sanitarias

Algoritmo para la realización de EE en ETS

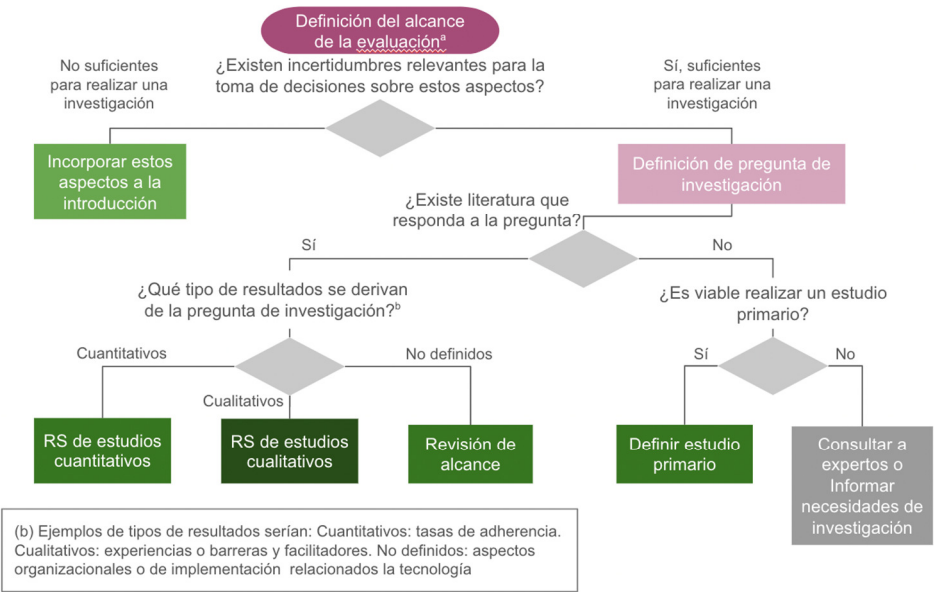
AIP: análisis de impacto presupuestario
 EE: evaluación económica
 ETS: evaluación de tecnologías sanitarias
 RS-EE: revisión sistemática de evaluaciones económicas
 TS: tecnología sanitaria

(a) ¿La TS es, al menos, no inferior al comparador?
 (b) EE aplicable: EE realizada en España, reciente, independiente y de calidad (valorar con escala de aplicabilidad, ej. NICE)
 (c) La viabilidad toma en cuenta la disponibilidad de información mínima y de recursos humanos, informáticos, etc.
 (d) EE:
 - EE completa - análisis de coste-efectividad, coste-utildad o coste-minimización
 - EE parcial - estudio de costes



Anexo 3. Algoritmo para la realización de los aspectos éticos, legales, organizativos, sociales y ambientales en Evaluación de Tecnologías Sanitarias

**ALGORITMO
PARA LA EVALUACIÓN
DE ASPECTOS ÉTICOS,
LEGALES,
ORGANIZATIVOS,
SOCIALES Y
AMBIENTALES**



[1] EUnetHTA Joint Action 2 Work Package 8. HTA Core Model © Version 3.0. 2016.
[2] Möberg J, Oxman AD, Rosenbaum S, Schünemann HJ, Guyatt G, Flottorp S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. Health Res. Policy Syst. 2018 may 29;16(1):45.
[3] The VALIDATE handbook: An approach on the integration of values in doing assessments of health technologies. Radboud University Press; 2022

Anexo 4. Listado de preguntas a los actores de interés

LISTADO DE PREGUNTAS REALIZADAS DE LAS PARTES INTERESADAS (PACIENTES, PROFESIONALES E INDUSTRIA)	
1.	¿Cuáles son los riesgos y beneficios que, desde su perspectiva, puede tener el uso de los sistemas automatizados para la detección de RAM?
2.	¿Para qué perfil de pacientes es deseable introducir esta tecnología? ¿Para cuáles no?
3.	¿Para qué problemas concretos puede ser adecuado el uso de esta tecnología?
LISTADO DE PREGUNTAS REALIZADAS A EXPERTOS/AS EN INFORMÁTICA	
CONTEXTO/GENERAL	
• ¿Qué experiencia tienen actualmente? ¿Qué sistema tienen en su hospital/área para detección de RAM? ¿Qué problemas creen que hay ahora, qué barreras tienen los sistemas actuales, etc.? • ¿Conocen algún sistema automatizado de detección de RAM que se aplique en su área? • ¿Se está planificando alguna implementación de algún sistema automatizado? ¿Cómo se está haciendo? • ¿Se prevén algunos problemas/ límites organizativos en esa implementación? ¿De qué tipo? • ¿Qué tipo de informáticos se ocupan del desarrollo e implantación de estos sistemas automatizados? Para el desarrollo/implantación de estos sistemas, hay coordinación entre los equipos informáticos y el resto de profesionales (equipos médicos, enfermeros y farmacéuticos)? • Hablando de estos sistemas automatizados, ¿sabrían decirnos de qué manera se puede evitar que haya una saturación de alarmas para los profesionales?	
ASPECTOS ÉTICOS, LEGALES, ORGANIZATIVOS Y SOCIALES	
• ¿Creen que se puede conseguir cierta interoperabilidad de los sistemas de detección de reacciones adversas entre diferentes hospitales, CC.AA.? ¿Sería viable que se creara un registro a nivel nacional de RAM?	

- ¿Cree que pueden darse problemas de privacidad y/o seguridad al implantar un sistema de detección automatizada de RAM por ejemplo entre CCAA?
- ¿Podrías decirnos cuál es la arquitectura tecnológica propuesta para un sistema de detección de RAM y cómo se integra con los sistemas de información existentes en el entorno clínico? ¿Qué problemas concretos o grado de complejidad cree que puede tener un sistema automatizado a la hora de implantarlo?
- Dada la rápida evolución de la IA, ¿podría decirse que sistemas desarrollados antes de una determinada fecha o periodo, a día de hoy estarían obsoletos con mucha probabilidad? Por ejemplo, lo hecho antes de 2015 seguramente haya sido ya superado en rendimiento. Una vez implantado, si a los dos años aparece otro sistema con un mejor rendimiento, ¿la sustitución es factible/viable/costosa?

EFFECTIVIDAD

- Un algoritmo o sistema de IA desarrollado para responder a las preguntas “¿Qué medicamentos producen cuáles efectos adversos? o ¿este medicamento produce tal efecto adverso, o produce efectos adversos no observados en los ensayos aleatorizados?” y que se aplica a grandes bases de datos de farmacovigilancia o de historias clínicas electrónicas multicéntricas, ¿es del mismo tipo o es directamente aplicable al que necesito para, a partir de la historia clínica de mis pacientes, responder a la pregunta “¿mis pacientes están sufriendo efectos adversos que están pasando desapercibidos como tales, o simplemente desapercibidos?”
- ¿Un sistema para detectar efectos adversos ya ocurridos, a partir de signos/síntomas y medicamentos en historia clínica, es distinto, en su arquitectura/programación, de un sistema que pretende predecir la aparición de efectos adversos a partir de las características del paciente y de sus medicamentos?
- ¿Un sistema de procesamiento del lenguaje natural (NLP) desarrollado para otro idioma es aprovechable en alguna medida para desarrollar uno en España, o la lengua específica obliga a desarrollarlo desde cero?

USO DE RECURSOS Y COSTES

- ¿Conocen o podrían conocer el coste que supone la implantación de este tipo de sistemas? Además del coste de implantación del sistema, ¿existe algún coste de mantenimiento?

CC.AA.: Comunidad Autónoma; IA: Inteligencia artificial; NPL: procesamiento del lenguaje natural; RAM: Reacción adversa a medicamentos;

Anexo 5. Preguntas sobre aspectos ambientales realizadas a la industria

Preguntas realizadas a empresa Costaisa
• ¿Conoce el impacto ambiental del desarrollo de los softwares para la implementación de sistemas automatizados de detección personalizados de reacciones adversas a medicamentos (RAM) basados en IA? • ¿Conocen el impacto ambiental (en términos de consumo eléctrico y huella de carbono) del mantenimiento de los sistemas automatizados de detección personalizados de RAM basados en IA? • Dada la rápida evolución de la Inteligencia Artificial (IA), ¿podría decirse que sistemas desarrollados antes de una determinada fecha o periodo, a día de hoy estarían obsoletos con mucha probabilidad? • Una vez implantado, si a los dos años aparece otro sistema con un mejor rendimiento, ¿la sustitución o renovación es factible/viable/costosa? • ¿Cómo se puede minimizar la fatiga por alertas que puedan tener los/as profesionales? ¿Existe gradación en el nivel de alertas? • ¿Conocen o podrían conocer el coste que supone la implantación de este tipo de sistemas? Además del coste de implantación del sistema, ¿existe algún coste de mantenimiento? • ¿Conoce el impacto ambiental del desarrollo de los softwares para la implementación de sistemas automatizados de detección personalizados de RAM basados en IA? • ¿Conocen el impacto ambiental (en términos de consumo eléctrico y huella de carbono) del mantenimiento de los sistemas automatizados de detección personalizados de RAM basados en IA?
RAM: Reacciones adversas a medicamentos; IA: Inteligencia artificial Consulta a la empresa Costaisa: Fecha de consulta: 02/11/2023

Anexo 6. Estimación fecha de actualización del informe



Anexo 7. Estudios excluidos a texto completo

Seguridad y efectividad

Excluidas por intervención (k=6)
- Hoppe C, Obermeier P, Muehlhans S, Alchikh M, Seeber L, Tief F, et al. Innovative Digital Tools and Surveillance Systems for the Timely Detection of Adverse Events at the Point of Care: A Proof-of-Concept Study. Drug Saf. 2016 oct;39(10):977-88. - Huang Y-L, Moon J, Segal JB. A Comparison of Active Adverse Event Surveillance Systems Worldwide. Drug Saf. 2014 ago;37(8):581-96. - Long J, Yuan MJ, Poonawala R. An Observational Study to Evaluate the Usability and Intent to Adopt an Artificial Intelligence–Powered Medication Reconciliation Tool. Interact J Med Res. 2016 may 16;5(2):e14. - Naor GM, Tocut M, Moalem M, Engel A, Feinberg I, Stein GY, et al. Screening for Medication Errors and Adverse Events Using Outlier Detection Screening Algorithms in an Inpatient Setting. J Med Syst. 2022 oct 26;46(12):88. - Noorda NMF, Sallevelt BTGM, Langendijk WL, Egberts TCG, Van Puijenbroek EP, Wilting I, et al. Performance of a trigger tool for detecting adverse drug reactions in patients with polypharmacy acutely admitted to the geriatric ward. Eur Geriatr Med. 2022 may 30;13(4):837-47. - Usui M, Aramaki E, Iwao T, Wakamiya S, Sakamoto T, Mochizuki M. Extraction and Standardization of Patient Complaints from Electronic Medication Histories for Pharmacovigilance: Natural Language Processing Analysis in Japanese. JMIR Med Inform. 2018 sep 27;6(3):e11021.
Excluidas por diseño (k=17)
- Arnaud M, Bégaud B, Thurin N, Moore N, Pariente A, Salvo F. Methods for safety signal detection in healthcare databases: a literature review. Expert Opinion on Drug Safety. 2017 jun 3;16(6):721-32. - Coste A, Wong A, Bokern M, Bate A, Douglas IJ. Methods for drug safety signal detection using routinely collected observational electronic health care data: A systematic review. Pharmacoepidemiology and Drug. 2023 ene;32(1):28-43. - Dubey SA, Pandit AA. A Comprehensive Review and Application of Interpretable Deep Learning Model for ADR Prediction. IJACSA. 2022;13(9). - Feng C, Le D, McCoy AB. Using Electronic Health Records to Identify Adverse Drug Events in Ambulatory Care: A Systematic Review. Appl Clin Inform. 2019 ene;10(1):123-8. - Injeti, N., Pal, A. Adverse Drug Reaction Monitoring: Pharmacovigilance and Artificial Intelligence. IJPR. 2021 feb 2;13(02). - Kaas - Hansen BS, Gentile S, Caioli A, Andersen SE. Exploratory pharmacovigilance with machine learning in big patient data: A focused scoping review. Basic Clin Pharma Tox. 2023 mar;132(3):233-41. - Karimi S, Wang C, Metke-Jimenez A, Gaire R, Paris C. Text and Data Mining Techniques in Adverse Drug Reaction Detection. ACM Comput. Surv. 2015 jul 21;47(4):1-39. - Kim HR, Sung M, Park JA, Jeong K, Kim HH, Lee S, et al. Analyzing adverse drug reaction using statistical and machine learning methods: A systematic review. Medicine. 2022 jun 24;101(25):e29387. - Liu F, Jagannatha A, Yu H. Towards Drug Safety Surveillance and Pharmacovigilance: Current Progress in Detecting Medication and Adverse Drug Events from Electronic Health Records. Drug Saf. 2019 ene;42(1):95-7.

Excluidas por diseño (k=17)

- Marien S, Krug B, Spinewine A. Electronic tools to support medication reconciliation: a systematic review. *Journal of the American Medical Informatics Association*. 2017 ene 1;24(1):227-40.
- On J, Park H-A, Yoo S. Development of a prediction models for chemotherapy-induced adverse drug reactions: A retrospective observational study using electronic health records. *European Journal of Oncology Nursing*. 2022 feb;56:102066.
- Ramos SF, Alvarez NR, Dos Santos Alcântara T, Sanchez JM, Da Costa Lima E, De Lyra Júnior DP. Methods for the detection of adverse drug reactions in hospitalized children: a systematic review. *Expert Opinion on Drug Safety*. 2021 oct 3;20(10):1225-36.
- Rawat A, Wani MA, ElAffendi M, Imran AS, Kastrati Z, Daudpota SM. Drug Adverse Event Detection Using Text-Based Convolutional Neural Networks (TextCNN) Technique. *Electronics*. 2022 oct 17;11(20):3336.
- Salas M, Petracek J, Yalamanchili P, Aimer O, Kasthuril D, Dhingra S, et al. The Use of Artificial Intelligence in Pharmacovigilance: A Systematic Review of the Literature. *Pharm Med*. 2022 oct;36(5):295-306.
- Shimai Y, Takeda T, Okada K, Manabe S, Teramoto K, Mihara N, et al. Screening of anticancer drugs to detect drug - induced interstitial pneumonia using the accumulated data in the electronic medical record. *Pharmacology Res & Perspec*. 2018 jul;6(4):e00421.
- Yung IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. *International Journal of Medical Informatics*. 2019 dic;132:103971.
- Zhao J, Henriksson A, Kvist M, Asker L, Boström H. Handling Temporality of Clinical Events for Drug Safety Surveillance. *AMIA Annu Symp Proc*. 2015;2015:1371-80.

Excluidas por objetivo (k=3)

- Combi C, Zorzi M, Pozzani G, Moretti U, Arzenton E. From narrative descriptions to MedDRA: automatically encoding adverse drug reactions. *Journal of Biomedical Informatics*. 2018 ago;84:184-99.
- Neue A, Wimmer S, Neubert A, Becker L, Prokosch H-U, Beckmann MW, et al. Towards a Computable Data Corpus of Temporal Correlations between Drug Administration and Lab Value Changes. *PloS ONE*. 2015 ago 24;10(8):e0136131.
- Shimai Y, Takeda T, Manabe S, Teramoto K, Mihara N, Matsumura Y. Method for Detecting Drug-Induced Interstitial Pneumonia from Accumulated Medical Record Data at a Hospital [Internet]. En: Chen Y-W, Torro C, Tanaka S, Howlett RJ, C. Jain L, 138reviam. *Innovation in Medicine and Healthcare 2015*. Cham: Springer International Publishing; 2016 [Accedido 26-12-2023]. P. 15-24.

Excluidas por medida de resultado (k=41)

- Alfattni G, Belousov M, Peek N, Nenadic G. Extracting Drug Names and Associated Attributes From Discharge Summaries: Text Mining Study. *JMIR Med Inform*. 2021 may 5;9(5):e24678.
- Bean DM, Wu H, Iqbal E, Dzahini O, Ibrahim ZM, Broadbent M, et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. *Sci Rep*. 2017 nov 27;7(1):16416.
- Chaichulee S, Promchai C, Kaewkomon T, Kongkamol C, Ingviya T, Sangsupawanich P. Multi-label classification of symptom terms from free-text bilingual adverse drug reaction reports using natural language processing. *PloS ONE*. 2022 ago 4;17(8):e0270595.
- Choudhury O, Park Y, Saloniadis T, Gkoulalas-Divanis A, Sylla I, Das AK. Predicting Adverse Drug Reactions on Distributed Health Data using Federated Learning. *AMIA Annu Symp Proc*. 2019;2019:313-22.

- Christopoulou F, Tran TT, Sahu SK, Miwa M, Ananiadou S. Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods. *Journal of the American Medical Informatics Association*. 2020 ene 1;27(1):39-46.
- Chu J, Dong W, He K, Duan H, Huang Z. Using neural attention networks to detect adverse medical events from electronic health records. *Journal of Biomedical Informatics*. 2018 nov;87:118-30.
- Dandala B, Joopudi V, Devarakonda M. Adverse Drug Events Detection in Clinical Notes by Jointly Modeling Entities and Relations Using Neural Networks. *Drug Saf*. 2019 ene;42(1):135-46.
- Dijkstra L, Schink T, Linder R, Schwaninger M, Pigeot I, Wright M, et al. A Discovery and Verification Approach for Pharmacovigilance using Electronic Health Care Data [Internet]. *Epidemiology*; 2022 [Accedido 22-12-2023]. Disponible en: <http://medrxiv.org/lookup/doi/10.1101/2022.05.10.22274885>
- Eriksson R, Werge T, Jensen LJ, Brunak S. Dose-Specific Adverse Drug Reaction Identification in Electronic Patient Records: Temporal Data Mining in an Inpatient Psychiatric Population. *Drug Saf*. 2014 abr;37(4):237-47.
- Geva A, Stedman JP, Manzi SF, Lin C, Savova GK, Avillach P, et al. Adverse drug event presentation and tracking (ADEPT): semiautomated, high throughput pharmacovigilance using real-world data. *JAMIA Open*. 2020 oct 1;3(3):413-21.
- Guan Y, Ji L, Zheng L, Yang J, Qin Y, Ding N, et al. Development of a drug risk analysis and assessment system and its application in signal excavation and analysis of 263 cases of fluoroquinolone-induced adverse reactions. *Front. Pharmacol*. 2022 oct 4;13:892503.
- Hazlehurst B, Green CA, Perrin NA, Brandes J, Carrell DS, Baer A, et al. Using natural language processing of clinical text to enhance identification of opioid - related overdoses in electronic health records data. *Pharmacoepidemiology and Drug*. 2019 ago;28(8):1143-51.
- Hoang T, Liu J, Roughead E, Pratt N, Li J. Supervised signal detection for adverse drug reactions in medication dispensing data. *Computer Methods and Programs in Biomedicine*. 2018 jul;161:25-38.
- Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Inform Assoc*. 2014 mar;21(2):353-62.
- Jagannatha A, Liu F, Liu W, Yu H. Overview of the First Natural Language Processing Challenge for Extracting Medication, Indication, and Adverse Drug Events from Electronic Health Record Notes (MADE 1.0). *Drug Saf*. 2019 ene;42(1):99-111.
- Jeong E, Park N, Choi Y, Park RW, Yoon D. Machine learning model combining features from algorithms with different analytical methodologies to detect laboratory-event-related adverse drug reaction signals. *PloS ONE*. 2018 nov 21;13(11):e0207749.
- Karlsson I, Bostrom H. Predicting Adverse Drug Events Using Heterogeneous Event Sequences [Internet]. En: 2016 IEEE International Conference on Healthcare Informatics (ICHI). Chicago, IL, USA: IEEE; 2016 [Accedido 22-12-2023]. P. 356-62.
- Kirkendall ES, Kouril M, Dexheimer JW, Courter JD, Hagedorn P, Szczesniak R, et al. Automated identification of antibiotic overdoses and adverse drug events via analysis of prescribing alerts and medication administration records. *Journal of the American Medical Informatics Association*. 2017 mar 1;24(2):295-302.
- Ledieu T, Bouzillé G, Thiessard F, Berquet K, Van Hille P, Renault E, et al. Timeline representation of clinical data: usability and added value for pharmacovigilance. *BMC Med Inform Decis Mak*. 2018 dic;18(1):86.
- Lee S, Choi J, Kim H-S, Kim GJ, Lee KH, Park CH, et al. Standard-based comprehensive detection of adverse drug reaction signals from nursing statements and laboratory results in electronic health records. *Journal of the American Medical Informatics Association*. 2017 jul 1;24(4):697-708.
- Létinier L, Jouganous J, Benkebil M, Bel - Létoile A, Goehrs C, Singier A, et al. Artificial Intelligence for Unstructured Healthcare Data: Application to Coding of Patient Reporting of Adverse Drug Reactions. *Clin Pharma and Therapeutics*. 2021 ago;110(2):392-400.
- Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. *J Am Med Inform Assoc*. 2014 mar;21(2):308-14.
- Ni Y, Lingren T, Hall ES, Leonard M, Melton K, Kirkendall ES. Designing and evaluating an automated system for real-time medication administration error detection in a neonatal intensive care unit. *Journal of the American Medical Informatics Association*. 2018 may 1;25(5):555-63.

- Niedrig DF, Hoppe L, Mächler S, Russmann H, Russmann S. Benzodiazepine Use During Hospitalization: Automated Identification of Potential Medication Errors and Systematic Assessment of Preventable Adverse Events. *PLoS ONE*. 2016 oct 6;11(10):e0163224.
- Pacurariu AC, Straus SM, Trifirò G, Schuemie MJ, Gini R, Herings R, et al. Useful Interplay Between Spontaneous ADR Reports and Electronic Healthcare Records in Signal Detection. *Drug Saf*. 2015 dic;38(12):1201-10.
- Parikh S, Christensen D, Stuchbery P, Peterson J, Hutchinson A, Jackson T. Exploring in-hospital adverse drug events using ICD-10 codes. *Aust. Health Review*. 2014;38(4):454.
- Park J, Lee S, Kim K, Jung J, Lee D. Large-scale prediction of adverse drug reactions-related proteins with network embedding. *Bioinformatics*. 2023 ene 1;39(1):btac843.
- Patadia VK, Schuemie MJ, Coloma P, Herings R, Van Der Lei J, Straus S, et al. Evaluating performance of electronic healthcare records and spontaneous reporting data in drug safety signal detection. *Int J Clin Pharm*. 2015 feb;37(1):94-104.
- Qin W, Zeng X, Jia Z, Zheng X, Duan H, Lu X, et al. Detecting Adverse Drug Event Signals from a Clinical CaseBase. *Stud Health Technol Inform*. 2017;245:549-53.
- Rozenblum R, Rodríguez-Monguio R, Volk LA, Forsythe KJ, Myers S, McGurrin M, et al. Using a Machine Learning System to Identify and Prevent Medication Prescribing Errors: A Clinical and Cost Analysis Evaluation. *The Joint Commission Journal on Quality and Patient Safety*. 2020 ene;46(1):3-10.
- Santiso S, Pérez A, Casillas A, Oronoz M. Neural negated entity recognition in Spanish electronic health records. *Journal of Biomedical Informatics*. 2020 may;105:103419.
- Shimai Y, Takeda T, Manabe S, Teramoto K, Mihara N, Matsumura Y. Development of a database and processing method for detecting hematotoxicity adverse drug events. *Stud Health Technol Inform*. 2015;210:65-9.
- Suárez-Paniagua V, Rivera Zavala RM, Segura-Bedmar I, Martínez P. A two-stage deep learning approach for extracting entities and relationships from medical texts. *Journal of Biomedical Informatics*. 2019 nov;99:103285.
- Sydykov, S.B.; Shopabaeva, A.R.; Zhakipbekov, K.S. et al. A centralized automated system of monitoring adverse events of medications- "Medreminder". 2018.
- Tham MY, Ye Q, Ang PS, Fan LY, Yoon D, Park RW, et al. Application and 140revisión140 of the Comparison on Extreme Laboratory Tests (CERT) algorithm for detection of adverse drug reactions: Transferability across national boundaries. *Pharmacoeconomics and Drug*. 2018 ene;27(1):87-94.
- Wang G, Jung K, Winnenburg R, Shah NH. A method for systematic discovery of adverse drug events from clinical notes. *Journal of the American Medical Informatics Association*. 2015 nov 1;22(6):1196-204.
- Wong A, Plasek JM, Montecalvo SP, Zhou L. Natural Language Processing and Its Implications for the Future of Medication Safety: A Narrative Review of Recent Advances and Challenges. *Pharmacotherapy*. 2018 ago;38(8):822-41.
- Wu T, Gao C-C, Lin J-S, Zha J-L. Active Monitoring of Adverse Drug Reactions with Neural Network Technology. *Chinese Medical Journal*. 2017 jun 20;130(12):1498-501.
- Yang X, Bian J, Gong Y, Hogan WR, Wu Y. MADEX: A System for Detecting Medications, Adverse Drug Events, and Their Relations from Clinical Notes. *Drug Saf*. 2019 ene;42(1):123-33.
- Zhan C, Roughead E, Liu L, Pratt N, Li J. A data-driven method to detect adverse drug events from prescription data. *Journal of Biomedical Informatics*. 2018 sep;85:10-20.
- Zhang W, Kuang Z, Peissig P, Page D. Adverse drug reaction discovery from electronic health records with deep neural networks [Internet]. En: *Proceedings of the ACM Conference on Health, Inference, and Learning*. Toronto Ontario Canada: ACM; 2020 [Accedido 22-12-2023]. P. 30-9.

Excluidas por tipo de publicación (k=10)

- Abril, G.M.; Garicano, M.V.; Campodónico, D.; Garcés, R.P.; Santos, F.A. Detection of adverse drug reactions in hospitalized patients through laboratory alerts: Hyponatremia and rhabdomyolysis. 2022. SUPPL 2. 130. Basic Clin. Pharmacol. Toxicol.
- Barman, H.; Venkateswaran, S.; Del Santo, A.; Rao, K.; Raghunatha, B.; Yoo, U.; Kottschade, L.; Block, M.; Scott Chandler, G.; Wagner, T.; Mohindra, R. IDENTIFICATION AND CHARACTERIZATION OF IMMUNE CHECKPOINT INHIBITOR INDUCED IMMUNE-RELATED ADVERSE EVENTS FROM ELECTRONIC HEALTH RECORDS USING NATURAL LANGUAGE PROCESSING TECHNIQUES. 2022.
- Iivanainen SME, Ekstrom J, Kataja V, Virtanen H, Koivunen J. 1841P Predicting the onset of immune-related adverse events (irAEs) in immune checkpoint inhibitor (ICI) therapies using a machine learning (ML) model trained with electronic patient-reported outcomes (ePROs) and lab measurements. Annals of Oncology. 2020 sep;31:S1057.
- Islam T, Hussain N, Islam S, Chakrabarty A. Detecting Adverse Drug Reaction with Data Mining And Predicting its Severity With Machine Learning [Internet]. En: 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). Malambe, Sri Lanka: IEEE; 2018 [Accedido 22-12-2023]. P. 1-5.
- Kaas-Hansen BS, Gentile S, Caioli A, Andersen SE. Exploratory pharmacovigilance with machine learning in big patient data: A focused scoping review. Basic Clin Pharma Tox. 2023 mar;132(3):233-41.
- Keith C, Martin J, Linger M. Letter to the Editor: Automated adverse drug reaction detection. Aust Prescr. 2018 oct 2;41(5):138.
- Lorenzini, K.I.; Foufi, V.; Tirone, F.; Rollason, V.; Gaudet-Blavignac, C.; Desmeules, J.; Lovis, C.; Samer, C. Screening of discharge letters to detect adverse drug reactions. Can natural language processing tools be helpful? 2018.
- Nugmanov, R.; Alimova, I.; Tutubalina, E. Adverse drug reactions identification in social media posts and electronic health records with neural networks. 2019.
- Truong, K.; Blackford, A.; Hoppes, J.; Odeh, A.; Patterson, T.; Schneck, S. Implementation of a clinical decision support system to guide appropriate medication use and monitoring in hospitalized patients with drug-induced QT interval prolongation. 2017.
- Xu W, Gusev A, Groha S, Rahma OE, Schrag D, Choueiri TK, et al. Automated identification of immune related adverse events in oncology patients using machine learning. JCO. 2021 may 20;39(15_suppl):1551-1551.

Excluidas por estar incluidos en otro estudio (k=15)

- Boyce R, Jao J, Miller T, Kane-Gill S. Automated Screening of Emergency Department Notes for Drug-Associated Bleeding Adverse Events Occurring in Older Adults. Appl Clin Inform. 2017;08(04):1022-30.
- Chapman AB, Peterson KS, Alba PR, DuVall SL, Patterson OV. Detecting Adverse Drug Events with Rapidly Trained Classification Models. Drug Saf. 2019 ene;42(1):147-56.
- Dandala B, Joopudi V, Tsou C-H, Liang JJ, Suryanarayanan P. Extraction of Information Related to Drug Safety Surveillance From Electronic Health Record Notes: Joint Modeling of Entities and Relations Using Knowledge-Aware Neural Attentive Models. JMIR Med Inform. 2020 jul 10;8(7):e18417.
- Gaebel J, Kolter T, Arlt F, Denecke K. Extraction Of Adverse Events From Clinical Documents To Support Decision Making Using Semantic Preprocessing. Stud Health Technol Inform. 2015;216:1030.
- Gupta S, Belouali A, Shah NJ, Atkins MB, Madhavan S. Automated Identification of Patients With Immune-Related Adverse Events From Clinical Notes Using Word Embedding and Machine Learning. JCO Clinical Cancer Informatics. 2021 dic;5(5):541-9.
- Kim Y, Meystre SM. Ensemble method-based extraction of medication and related information from clinical texts. Journal of the American Medical Informatics Association. 2020 ene 1;27(1):31-8.

- Li F, Liu W, Yu H. Extraction of Information Related to Adverse Drug Events from Electronic Health Record Notes: Design of an End-to-End Model Based on Deep Learning. *JMIR Med Inform.* 2018 nov 26;6(4):e12159.
- Li F, Yu H. An investigation of single-domain and multidomain medication and adverse drug event relation extraction from electronic health record notes using advanced deep learning models. *Journal of the American Medical Informatics Association.* 2019 jul 1;26(7):646-54.
- Munkhdalai T, Liu F, Yu H. Clinical Relation Extraction Toward Drug Safety Surveillance Using Electronic Health Record Narratives: Classical Learning Versus Deep Learning. *JMIR Public Health Surveill.* 2018 abr 25;4(2):e29.
- Oronoz M, Gojenola K, Pérez A, De Ilarraza AD, Casillas A. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. *Journal of Biomedical Informatics.* 2015 ago;56:318-32.
- Rebane J, Samsten I, Papapetrou P. Exploiting complex medical data with interpretable deep learning for adverse drug event prediction. *Artificial Intelligence in Medicine.* 2020 sep;109:101942
- Santiso S, Casillas A, Pérez A. The class imbalance problem detecting adverse drug reactions in electronic health records. *Health Informatics J.* 2019 dic;25(4):1768-78.
- Santiso S, Perez A, Casillas A. Exploring Joint AB-LSTM With Embedded Lemmas for Adverse Drug Reaction Discovery. *IEEE J. Biomed. Health Inform.* 2019 sep;23(5):2148-55.
- Taewijit S, Theeramunkong T, Ikeda M. Distant Supervision with Transductive Learning for Adverse Drug Reaction Identification from Electronic Medical Records. *Journal of Healthcare Engineering.* 2017;2017:1-21.
- Wunnava S, Qin X, Kakar T, Sen C, Rundensteiner EA, Kong X. Adverse Drug Event Detection from Electronic Health Records Using Hierarchical Recurrent Neural Networks with Dual-Level Embedding. *Drug Saf.* 2019 ene;42(1):113-22.

Aspectos éticos, legales, organizativos, sociales y ambientales

Excluidas por intervención	Motivo Exclusión
Chen B, Restaino J, Tippet E. Key Elements in Adverse Drug Reactions Safety Signals: Application of Legal Strategies. Cancer Treat Res. 2019;171:47-59. Doi: 10.1007/978-3-319-43896-2_4.	No aborda sistemas automatizados de detección de RAM, y se centra en aspectos legales de la farmacovigilancia en general en el contexto estadounidense
Salathé M. Digital Pharmacovigilance and Disease Surveillance: Combining Traditional and Big-Data Systems for Better Public Health. J Infect Dis. 2016 Dec 1;214(suppl_4):S399-S403. Doi: 10.1093/infdis/jiw281. PMID: 28830106	No aborda sistemas automatizados de detección de RAM, y se centra en otros aspectos de la farmacovigilancia
Jafarpour B, Abidi SR, Ahmad AM, Abidi SS. INITIATE: An Intelligent Adaptive Alert Environment. Stud Health Technol Inform. 2015;216:285-9.	No aborda sistemas automatizados de detección de RAM, y se trata el tema de las alarmas en general
Ramalingam S, Ponnuswamy TK, Sivan YS. Innovations in monitoring of adverse drug reactions: the role of a technical advisor. Indian J Med Ethics. 2013 Jul-Sep;10(3):190-1. Doi: 10.20529/IJME.2013.055	Aborda la monitorización de RAM durante los ensayos clínicos, no en el periodo postmarketing
Härmark L, van Grootheest AC. Pharmacovigilance: methods, recent developments and future perspectives. Eur J Clin Pharmacol. 2008 Aug;64(8):743-52. Doi: 10.1007/s00228-008-0475-9.	No aborda sistemas automatizados de detección de RAM, y se centra en otros aspectos de la farmacovigilancia

Excluidas por Diseño	Motivo Exclusión
Skalafouris C, Reny JL, Stirnemann J, Groscurin O, Eggimann F, Grauser D, Teixeira D, Jermini M, Bruggmann C, Bonnabry P, Guignard B. Development and assessment of PharmaCheck: an electronic screening tool for the prevention of twenty major adverse drug events. BMC Med Inform Decis Mak. 2022 May 31;22(1):146. Doi: 10.1186/s12911-022-01885-8.	Sólo disponible abstract
Carroll AE. Averting Alert Fatigue to Prevent Adverse Drug Reactions. JAMA. 2019 Aug 20;322(7):601. Doi: 10.1001/jama.2019.11710	Sólo disponible abstract
Rush JL, Ibrahim J, Saul K, Brodell RT. Improving Patient Safety by Combating Alert Fatigue. J Grad Med Educ. 2016 Oct;8(4):620-621. Doi: 10.4300/JGME-D-16-00186.1.	Carta al editor
Pai S, Munshi R. Real life challenges in implementing the Pharmacovigilance Programme of India (PvPI)-Insights from an adverse drug reaction monitoring centre. Pharma Times, 2015	Sólo disponible abstract
Hopf, Y.; Bond, C.; Haughney, J.; Helms, P. Data linkage for pharmacovigilance – Seeking the views and opinions of healthcare professionals. International Journal of Pharmacy Practice, 2011	Sólo disponible abstract
Cheng CM. Hospital systems for the detection and prevention of adverse drug events. Clin Pharmacol Ther. 2011 Jun;89(6):779-81. Doi: 10.1038/clpt.2010.356	Artículo de opinión
Wood, S.M. Adverse Drug Reactions Online Information Tracking (ADROIT): Development of a new computer system to support adverse drug reaction monitoring. Pharmaceutical Medicine, 1989	Sólo disponible abstract

Excluidas por alcance	Motivo Exclusión
Murphy RM, Klopotoska JE, de Keizer NF, Jager KJ, Leopold JH, Dongelmans DA, Abu-Hanna A, Schut MC. Adverse drug event detection using natural language processing: A scoping review of supervised learning methods. PloS One. 2023 Jan 3;18(1):e0279842. Doi: 10.1371/journal.pone.0279842.	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Yasrebi-de Kom IAR, Dongelmans DA, de Keizer NF, Jager KJ, Schut MC, Abu-Hanna A, Klopotoska JE. Electronic health record-based prediction models for in-hospital adverse drug event diagnosis or prognosis: a systematic review. J Am Med Inform Assoc. 2023 Apr 19;30(5):978-988. Doi: 10.1093/jamia/ocad014.	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Skalafouris C, Reny JL, Stirnemann J, Grosgrin O, Eggimann F, Grauser D, Teixeira D, Jermini M, Bruggmann C, Bonnabry P, Guignard B. Development and assessment of PharmaCheck: an electronic screening tool for the prevention of twenty major adverse drug events. BMC Med Inform Decis Mak. 2022 May 31;22(1):146. Doi: 10.1186/s12911-022-01885-8.	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Wadhwa D, Kumar K, Batra S, Sharma S. Automation in signal management in pharmacovigilance-an insight. Brief Bioinform. 2021 Jul 20;22(4):bbaa363. Doi: 10.1093/bib/bbaa363.	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Nguyen DA, Nguyen CH, Mamitsuka H. A survey on adverse drug reaction studies: data, tasks and machine learning methods. Brief Bioinform. 2021 Jan 18;22(1):164-177. Doi: 10.1093/bib/bbz140.	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Ramos SF, Alvarez NR, Dos Santos Alcântara T, Sanchez JM, da Costa Lima E, de Lyra Júnior DP. Methods for the detection of adverse drug reactions in hospitalized children: a systematic review. Expert Opin Drug Saf. 2021 Oct;20(10):1225-1236. Doi: 10.1080/14740338.2021.1924668	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Choudhury O, Park Y, Salonidis T, Gkoulalas-Divanis A, Sylla I, Das AK. Predicting Adverse Drug Reactions on Distributed Health Data using Federated Learning. AMIA Annu Symp Proc. 2020 Mar 4;2019:313-322.	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Combi C, Zorzi M, Pozzani G, Moretti U, Arzenton E. From narrative descriptions to MedDRA: automatically encoding adverse drug reactions. J Biomed Inform. 2018 Aug;84:184-199. Doi: 10.1016/j.jbi.2018.07.001.	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Luo Y, Thompson WK, Herr TM, Zeng Z, Berendsen MA, Jonnalagadda SR, Carson MB, Starren J. Natural Language Processing for HER-Based Pharmacovigilance: A Structured Review. Drug Saf. 2017 Nov;40(11):1075-1089. Doi: 10.1007/s40264-017-0558-6.	No aborda aspectos éticos, legales, organizativos, sociales o ambientales
Carroll AE. Averting Alert Fatigue to Prevent Adverse Drug Reactions. JAMA. 2019 Aug 20;322(7):601. Doi: 10.1001/jama.2019.11710	Sólo disponible abstract
Rush JL, Ibrahim J, Saul K, Brodell RT. Improving Patient Safety by Combating Alert Fatigue. J Grad Med Educ. 2016 Oct;8(4):620-621. Doi: 10.4300/JGME-D-16-00186.1.	Carta al editor
Pai S, Munshi R. Real life challenges in implementing the Pharmacovigilance Programme of India (PvPI)-Insights from an adverse drug reaction monitoring centre. Pharma Times, 2015	Sólo disponible abstract

Hopf, Y.; Bond, C.; Haughney, J.; Helms, P. Data linkage for pharmacovigilance – Seeking the views and opinions of healthcare professionals. International Journal of Pharmacy Practice, 2011	Sólo disponible abstract
Cheng CM. Hospital systems for the detection and prevention of adverse drug events. Clin Pharmacol Ther. 2011 Jun;89(6):779-81. Doi: 10.1038/clpt.2010.356	Artículo de opinión
Wood, S.M. Adverse Drug Reactions Online Information Tracking (ADROIT): Development of a new computer system to support adverse drug reaction monitoring. Pharmaceutical Medicine, 1989	Sólo disponible abstract
Excluidas por idioma	Motivo Exclusión
Choi NK, Park BJ. [Adverse drug reaction surveillance system in Korea]. J Prev Med Public Health. 2007 Jul;40(4):278-84. Korean. Doi: 10.3961/jpmph.2007.40.4.278.	Idioma coreano

Anexo 8. Características de los estudios incluidos (aspectos éticos, legales, organizativos, sociales y ambientales)

Características principales de los estudios incluidos								
Estudio	País	Objetivo	Diseño	Población	Hallazgos principales	Limitaciones metodológicas	Instrumento medición de calidad	Conflicto de interés
Amin, 2016	Pakistán	Resumir los resultados de investigaciones realizadas en países desarrollados y en vías de desarrollo, así como en Pakistán, sobre la evolución de las tendencias y las perspectivas de futuro del seguimiento y la notificación de las RAM por parte de farmacéuticos/as.	Revisión Narrativa	N/A	Los programas de notificación de RAM que involucran a los equipos médicos, de enfermería y farmacéuticos son más fructíferos en la gestión y la notificación en comparación con aquellos que involucran únicamente a equipos médicos. Los sistemas de seguimiento dirigidos por hospitales para informar e identificar las RAM también se pueden utilizar como complemento de otros sistemas de notificación espontánea para mejorar el seguimiento de las RAM	No limitaciones destacables	SANRA	No reportado
Chen, 2021	Estados Unidos	Analizar diferentes métodos de protección de datos en los sistemas de farmacovigilancia	Revisión Narrativa	N/A	Aunque se ha demostrado que la aplicación del enfoque de "grandes datos" ayuda a detectar RAM al involucrar fuentes de datos que antes no estaban disponibles y promover el intercambio de información de salud, aún persisten desafíos y riesgos potenciales. La falta de	No limitaciones destacables	SANRA	Declaran no tener conflicto de interés

Características principales de los estudios incluidos								
Estudio	País	Objetivo	Diseño	Población	Hallazgos principales	Limitaciones metodológicas	Instrumento medición de calidad	Conflicto de interés
					medidas efectivas para preservar la privacidad en el flujo de datos médicos es la más importante, por lo que se requieren acciones urgentes para prevenir las amenazas y facilitar la construcción de sistemas de farmacovigilancia.			
Edrees, 2022	Estados Unidos	Se analizan dos casos en los que los métodos de inteligencia artificial pueden ser útiles para mejorar la calidad de la farmacovigilancia y el papel de la inteligencia artificial en las prácticas de telesalud.	Revisión Narrativa	N/A	La IA en telesalud se puede utilizar para mejorar la farmacovigilancia mediante la utilización de diversas fuentes de información del paciente, como registros médicos electrónicos, tecnologías de información sanitaria y sistemas de bases de datos de farmacovigilancia, para detectar y prevenir problemas relacionados con los medicamentos. Aunque la IA tiene un papel prometedor en el campo de la farmacovigilancia y la telesalud, todavía existen desafíos que incluyen la detección de eventos adversos de medicamentos desconocidos o no documentados, preocupaciones sobre la privacidad y dificultades técnicas.	No limitaciones destacables	SANRA	DW.B informa subvenciones y honorarios personales (HP) de EarlySense, HP de CDI Negev, capital de ValeraHealth, capital de Clew, capital de MDClone, HP y capital de AESOP, HP y capital de Feelbetter y subvenciones de IBM Watson Health. Todos los demás autores no tienen

Características principales de los estudios incluidos								
Estudio	País	Objetivo	Diseño	Población	Hallazgos principales	Limitaciones metodológicas	Instrumento medición de calidad	Conflicto de interés
								conflictos de intereses que revelar.
Glassman, 2006	Estados Unidos	Examinar algunos de los efectos indirectos potenciales de la introducción informatizada de pedidos y de un sistema automatizado de alerta de medicamentos.	Estudio observacional mediante encuesta	Médicos/as de un hospital de veteranos/as (n=97)	La exposición a intervalos a alertas automáticas de medicamentos tuvo poco o ningún efecto en el reconocimiento de interacciones medicamentosas seleccionadas. La principal barrera percibida para la utilización eficaz de las alertas sobre drogas siguió siendo la misma a lo largo del tiempo.	Uso de datos auto-reportados, encuestas realizadas en un único sistema electrónico, por lo que puede no ser aplicable a otros entornos, el estudio es de 2006	MMAT	No reportado
Hurdle, 2003	Estados Unidos	Explorar numerosas RAM no detectadas (y por lo tanto invisible) en el HCE. Especulan cómo persisten estas narrativas invisibles de RAM de enfermería y qué presagian para la reingeniería de la seguridad.	Estudio retrospectivo mediante revisión de registros	Red de hospitales de veteranos/as	Observaron que los/as enfermeros/as detectaron sólo alrededor del 14% de los RAM en un estudio previo. Sin embargo, las notas de enfermería informales, a menudo no electrónicas, aumentaron la incidencia global de los RAM en casi un tercio. Enfermería es una fuente excelente de señales para los RAM, pero es fácil subestimar su papel en este sentido, especialmente en los estudios que se basan en la revisión "clásica" de las HCE, ya que se tiende a pasar por alto las anotaciones no normativas (por ejemplo, el texto garabateado en una hoja de flujo o que es un fragmento	No limitaciones destacables	N/A	No reportado

Características principales de los estudios incluidos								
Estudio	País	Objetivo	Diseño	Población	Hallazgos principales	Limitaciones metodológicas	Instrumento medición de calidad	Conflicto de interés
					escueto en un campo informático estrecho como “comentarios” en un registro electrónico de administración de medicación). Estos datos subrayan la necesidad de una reingeniería de las HCE para tomar en serio la comunicación informal y a menudo invisible que tiene lugar a nivel de enfermería.			
Li, 2022 ^a	Australia	Identificar factores asociados con la notificación de RAM y mapearlos en un marco de cambio de comportamiento para ayudar a informar futuras intervenciones diseñado para mejorar el subregistro de RAM.	Métodos Mixtos. Se agregó un componente cualitativo a un estudio principalmente cuantitativo y los datos se recopilaron y analizaron juntos.	Profesionales sanitarios/as (médicos/as, enfermeros/as y farmacéuticos/as hospitalarios/as (n=133)	Saber cómo notificar las RAM, haber recibido formación sobre la notificación de RAM y encontrarse con RAM como parte de la práctica clínica fueron predictores significativos de la notificación de una RAM. El análisis de contenido identificó tres categorías: modificar el proceso de notificación de RAM, permitir a los médicos informar las RAM y crear una cultura positiva de notificación de RAM. Después de la integración de los datos, los tres dominios objetivo del estudio fueron conocimiento, contexto/recursos ambientales y creencias sobre las consecuencias.	Muestra limitada tamaño completo de 133 profesionales de la salud en un solo hospital, puede haber sesgo de deseabilidad social, aunque el uso de encuestas anónimas puede haber reducido el impacto de este sesgo.	MMAT	Declaran no tener conflicto de interés
Li, 2022b	Australia	Investigar cómo aprovechar las aplicaciones	Revisión Narrativa	N/A	La falta de notificación de RAM sigue siendo un problema importante en los	No limitaciones destacables	SANRA	Declaran no tener

Características principales de los estudios incluidos								
Estudio	País	Objetivo	Diseño	Población	Hallazgos principales	Limitaciones metodológicas	Instrumento medición de calidad	Conflicto de interés
		digitales y la automatización en toda la industria de la salud mejorar la cantidad y calidad de los informes RAM.			sistemas de salud de todo el mundo, y múltiples estudios muestran que la magnitud de la falta de notificación fue consistentemente superior al 90% tanto en la atención primaria como en los entornos hospitalarios. A pesar de que los facilitadores y las barreras han sido identificados en la literatura, no han sido relacionados con un cambio de comportamiento, lo que es esencial para el diseño de una intervención eficaz. La mayoría de las intervenciones implementadas para mejorar la presentación de informes de ADR fueron modestamente exitosas y solo por un corto período, volviendo a los niveles previos a la intervención dentro de los 6 a 12 meses posteriores a la implementación.			conflicto de interés
Muzaffar 2023	Estados Unidos	Revisar el uso existente de IA para farmacovigilancia. Analizar dos casos de uso en los que estos métodos pueden ser útiles para mejorar la	Revisión Narrativa	N/A	Investigaciones recientes han identificado varios problemas con el uso de HCE para monitorizar RAM. Estos incluyen la falta de estandarización entre sistemas de HCE, especificidad en las opciones de ingreso de datos,	No limitaciones destacables	SANRA	Declaran no tener conflicto de interés

Características principales de los estudios incluidos								
Estudio	País	Objetivo	Diseño	Población	Hallazgos principales	Limitaciones metodológicas	Instrumento medición de calidad	Conflicto de interés
		calidad de la farmacovigilancia y el papel de la IA en telesalud.			documentación incompleta e inexacta, y fatiga alerta. Estos problemas pueden limitar la eficacia de la monitorización de RAM y comprometer la seguridad del paciente.			
Tagne, 2023	Australia	Explorar la baja tasa de notificación (barreras y facilitadores) de reacciones adversas por parte de los farmacéuticos comunitarios (FC) en Australia, evaluar la facilidad de uso de un sistema de notificación existente, y cómo este conocimiento puede influir en el diseño de posteriores sistemas electrónicos de notificación de reacciones adversas.	Métodos mixtos	Farmacéuticos comunitarios (n=12; n=7)	Los resultados muestran que los FC están dispuestos a informar las RAM, pero enfrentan barreras provenientes de infraestructuras ambientales, organizativas y técnicas. Aumentar la concienciación sobre las RAM, mejorar las prácticas en el lugar de trabajo y aplicar sistemas de presentación de informes electrónicos centrados en el usuario eran factores que facilitaban la presentación de informes de RAM. Los elementos de diseño, como un menú desplegable, entrada de texto libre, casillas de verificación y campos de datos precargados, se consideraron extremadamente útiles para navegar por el sistema y facilitar la presentación de informes RAM.	Se utilizó un muestreo intencional que podría haber dado lugar a un sesgo de selección. La generalización a otras áreas clínicas puede estar limitada ya que las infraestructuras de información, la organización del trabajo, la cultura o las condiciones y tareas ambientales varían.	MMAT	Declaran no tener conflicto de interés
Tillman, 2020	Estados Unidos	Evaluar los RAM dentro de los registros médicos electrónicos y	Estudio retrospectivo mediante	Área pediátrica	Varios factores pueden afectar la tasa de documentación RAM en emergencias pediátricas. La identificación	Se llevó a cabo en una sola institución, lo que limita	N/A	No reportado

Características principales de los estudios incluidos								
Estudio	País	Objetivo	Diseño	Población	Hallazgos principales	Limitaciones metodológicas	Instrumento medición de calidad	Conflicto de interés
		determinar los factores específicos asociados con la documentación de RAM apropiada y oportuna.	revisión de registros		de factores como medicamentos específicos, tipos de RAM y entornos de tratamiento clínico asociados con una menor documentación de RAM puede ayudar a identificar áreas específicas y brindar apoyo para mayores esfuerzos de farmacovigilancia.	la generalización. Debido a la gran cantidad de RAM informado, no se pudieron proporcionar datos específicos de pacientes sobre los resultados de las RAM.		

Anexo 9. Valoración de las limitaciones metodológicas de estudios incluidos (aspectos éticos, legales, organizativos, sociales y ambientales)

Valoración de las limitaciones metodológicas de las revisiones narrativas con el instrumento SANRA						
PREGUNTAS		Amin, 2016	Edress, 2022	Chen, 2021	Muzafar, 2023	Li, 2022.b
Justification of the article's importance for the readership		2	2	2	1	2
The importance is not justified	0					
The importance is alluded to, but not explicitly justified	1					
The importance is explicitly justified	2					
Statement of concrete aims or formulations of questions		2	2	1	1	2
No aims or questions are formulated	0					
Aims are formulated generally but not concretely or in terms of clear questions	1					
One or more concrete aims or questions are formulated 2	2					
Description of the literature search		2	0	0	0	2
The search strategy is not presented	0					
The literature search is described briefly	1					
The literature search is described in detail, including search terms and inclusion criteria	2					
Referencing		2	2	2	2	2
Key statements are not supported by references	0					

Valoración de las limitaciones metodológicas de las revisiones narrativas con el instrumento SANRA

PREGUNTAS		Amin, 2016	Edress, 2022	Chen, 2021	Muzafar, 2023	Li, 2022.b
The referencing of key statements is inconsistent	1					
Key statements are supported by references	2					
Scientific reasoning		2	2	2	2	2
The article's point is not based on appropriate arguments	0					
Appropriate evidence is introduced selectively	1					
Appropriate evidence is generally presented	2					
Appropriate presentation of data		2	2	2	2	2
Data are presented inadequately	0					
Data are often not presented in the most appropriate way	1					
Relevant outcome data are generally presented appropriately	2					
Puntuación total		12	10	9	8	12

Valoración de las limitaciones metodológicas de los estudios cualitativos con la lista de comprobación CASPe	
PREGUNTAS	Tagne, 2023
1. ¿Se definieron de forma clara los objetivos de la investigación?	Si
2. ¿Es congruente la metodología cualitativa?	Si
3. ¿El método de investigación es adecuado para alcanzar los objetivos?	Si
4. ¿La estrategia de selección de participantes es congruente con la pregunta de investigación y el método utilizado?	Si
5. ¿Las técnicas de recogida de datos utilizados son congruentes con la pregunta de investigación y el método utilizado?	Si
6. ¿Se ha reflexionado sobre la relación entre el investigador y el objeto de investigación (reflexibilidad)?	No
7. ¿Se han tenido en cuenta los aspectos éticos?	Si
8. ¿El análisis de los datos fue suficientemente riguroso?	Si
9. ¿Existe una exposición clara de los hallazgos?	Si
10. ¿Resulta de valor este estudio?	Si

Valoración de las limitaciones metodológicas con el instrumento MMAT				
	Methodological quality criteria	Raymond Li, 2022		
		Responses		
		YES	NO	CAN'T TELL
Screening questions (for all types)	S1. Are there clear research questions?	X		
	S2. Do the collected data allow to address the research questions?	X		
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.			
1. Qualitative	1.1. Is the qualitative approach appropriate to answer the research question?	X		
	1.2. Are the qualitative data collection methods adequate to address the research question?	X		
	1.3. Are the findings adequately derived from the data?	X		
	1.4. Is the interpretation of results sufficiently substantiated by data?	X		
	1.5. Is there coherence between qualitative data sources, collection, analysis and interpretation?	X		
2. Quantitative randomized controlled trials	2.1. Is randomization appropriately performed?			
	2.2. Are the groups comparable at baseline?			
	2.3. Are there complete outcome data?			
	2.4. Are outcome assessors blinded to the intervention provided?			
	2.5 Did the participants adhere to the assigned intervention?			

3. Quantitative non- randomized	3.1. Are the participants representative of the target population?			
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?			
	3.3. Are there complete outcome data?			
	3.4. Are the confounders accounted for in the design and analysis?			
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?			
4. Quantitative descriptive	4.1. Is the sampling strategy relevant to address the research question?			
	4.2. Is the sample representative of the target population?			
	4.3. Are the measurements appropriate?			
	4.4. Is the risk of nonresponse bias low?			
	4.5. Is the statistical analysis appropriate to answer the research question?			
5. Mixed methods	5.1. Is there an adequate rationale for using a mixed methods design to address the research question?	X		
	5.2. Are the different components of the study effectively integrated to answer the research question?	X		
	5.3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted?	X		
	5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?			X
	5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?	X		
Calidad total:		Alta		

Escala de valoración de la evidencia para series de casos (IHE)

Dominio	Tillman, 2020	Hurdle, 2003
1. ¿El objetivo del estudio se define claramente en el resumen, introducción o metodología?	Sí	Sí
2. ¿Se describen las características de los participantes?	No	No
3. Los casos incluidos, ¿proceden de más de un centro?	Sí	Sí
4. ¿Los criterios de elegibilidad (criterios de inclusión y exclusión) para entrar en el estudio son explícitos y apropiados?	Incierto	Incierto
5. ¿Los participantes fueron reclutados consecutivamente?	No	No
6. ¿Los participantes entraron en el estudio en la misma fase de la enfermedad?	Incierto	Incierto
7. ¿Se describe claramente la intervención en el estudio?	NA	NA
8. ¿Las intervenciones adicionales (co-intervenciones) fueron descritas claramente?	NA	NA
9. Las medidas de resultado, ¿son descritas en la introducción o el apartado de metodología?	Sí	Sí
10. Los resultados relevantes, ¿fueron medidos de forma apropiada con métodos objetivos y/o subjetivos?	Sí	Sí
11. ¿Los resultados fueron medidos antes y después de la intervención?	NA	NA
12. ¿Fueron apropiados los test estadísticos utilizados para evaluar los resultados relevantes?	Sí	Sí
13. ¿Se describe la duración del seguimiento?	NA	NA
14. ¿Se describen las pérdidas durante el seguimiento?	NA	NA
15. En el análisis de los resultados relevantes ¿Proporciona el estudio estimaciones de la variabilidad?	NA	NA
16. ¿Se describen los efectos adversos?	NA	NA
17. ¿Las conclusiones del estudio se basan en los resultados obtenidos?	Sí	Sí
18. ¿Se realiza una declaración de intereses y se describen las fuentes de financiación?	No	No
Número total de respuestas (+= sí; -=no; P=parcialmente; I=incierto)	Sí = 6 No = 3 Incierto = 2 No Aplica=7	Sí = 6 No = 3 Incierto = 2 No Aplica= 7
Riesgo general de sesgo	Baja calidad metodológica	Baja calidad metodológica

Fuente: tomado y traducido de: Institute of Health Economics (IHE). Quality Appraisal of Case Series Studies Checklist. Edmonton (AB): Institute of Health Economics; 2014. Disponible en: <http://www.ihe.ca/research-programs/rmd/cssqac/cssqac-about>

Anexo 10. Marco de los principios de la OMS para garantizar una IA responsable

El objetivo del desarrollo de una IA responsable (ver tabla a continuación) es evitar que las decisiones algorítmicas perpetúen y refuercen los prejuicios y desigualdades sociales que ya existen en nuestra sociedad y evitar una mayor discriminación. Uno de los mayores desafíos que enfrentan los/as científicos/as de datos al corregir el sesgo es la posible pérdida de calidad en un sistema de IA, ya que esta corrección puede significar una menor precisión del sistema. Sin embargo, es importante evitar sesgos injustos, ya que los conjuntos de datos utilizados por los sistemas de IA pueden contener sesgos históricos o modelos de gestión incorrectos, que pueden conducir a prejuicios y discriminación implícitos e inconscientes contra ciertos grupos o individuos, reforzando los estereotipos y marginación (Inteligencia Artificial Ética en Sanidad). Además de la “injusticia algorítmica”, existe el riesgo de que el uso de algoritmos pueda causar una impresión inapropiada de que son infalibles o neutrales éticamente, es decir, que no se basan en decisiones que incluyen prioridades de algunos valores en comparación con otras decisiones [89]. Por eso, es muy importante ser consciente de los problemas relacionados con los sesgos, saber detectarlos, reconocer sus causas y eliminarlos o, si no es posible, al menos minimizarlos. Si es posible, los algoritmos deberían ser auditados, con especial énfasis en su confiabilidad, trazabilidad y valor ético (Inteligencia Artificial Ética en Sanidad).

Marco de principios de la OMS para garantizar que la IA funcione para el interés público en todos los países.	
Proteger la autonomía humana	Las personas deben mantener el control de los sistemas de atención sanitaria y las decisiones médicas; la privacidad y la confidencialidad deben estar protegidas, y los/as pacientes deben dar un consentimiento informado válido a través de marcos legales apropiados para la protección de datos.
Promover el bienestar y la seguridad humanos y el interés público	Al diseñar tecnologías de IA se deben satisfacer los requisitos normativos de seguridad, precisión y eficacia para casos de uso o indicaciones bien definidos. Debe disponerse de medidas de control de la calidad en la práctica y de mejora de la calidad en el uso de la IA.
Garantizar la transparencia, la explicabilidad y la inteligibilidad.	La transparencia exige que se publique información suficiente antes del diseño o despliegue de una tecnología de IA. Esta información debe ser fácilmente accesible y facilitar una consulta y un debate públicos significativos sobre cómo se diseña la tecnología y cómo debe o no utilizarse.
Fomentar la responsabilidad y la rendición de cuentas.	Es responsabilidad de las partes interesadas garantizar que se utilicen los sistemas de IA en condiciones adecuadas y por personas debidamente formadas, deben existir mecanismos eficaces de cuestionamiento y reparación para las personas y grupos que se vean perjudicados por decisiones basadas en algoritmos.

Garantizar la inclusión y la equidad.	La IA para la salud debe diseñarse para fomentar un uso y un acceso lo más equitativo posible, independientemente de la edad, el sexo, el género, los ingresos, la raza, la etnia, la orientación sexual, la capacidad u otras características protegidas por los códigos de derechos humanos.
Promover una IA que responda a las necesidades y sea sostenible	Los/as diseñadores/as, desarrolladores/as y usuarios/as deben evaluar de forma continua y transparente las aplicaciones de IA durante su uso real para determinar si la IA responde adecuadamente a las expectativas. Los sistemas de IA también deben diseñarse para minimizar sus consecuencias medioambientales y aumentar la eficiencia energética. Los gobiernos y las empresas deben abordar los trastornos previstos en el lugar de trabajo, incluida la formación del personal sanitario para adaptarse al uso de los sistemas de IA, y las posibles pérdidas de puestos de trabajo debidas al uso de sistemas automatizados.

Anexo 11. Características de los estudios incluidos (Necesidades de investigación y medidas de resultado estándares)

Tabla 13. Principales características de los estudios incluidos. Necesidades de investigación y medidas de resultado estándares, en el uso de sistemas automatizados de detección de reacciones adversas a medicamentos			
Autor principal (año)	Objetivo	Diseño / Tipo de Estudio	Población
Núñez et al. (en marcha) [92]	Desarrollar un conjunto de resultados básicos para la seguridad del paciente en la atención perioperatoria.	Métodos mixtos: grupo focal, revisiones sistemáticas y metodología Delphi de consenso.	Expertos clínicos, metodólogos, investigadores, agencias gubernamentales, pacientes, cuidadores y representantes del paciente.
Jungreithmayr et al. (en marcha) [93]	Desarrollar un conjunto de resultados básicos para la evaluación de la implementación de un sistema de Entrada de Orden Médica Computarizada (CPOE).	Métodos mixtos: grupo focal, revisiones sistemáticas, entrevistas y metodología Delphi de consenso.	Expertos clínicos, investigadores, pacientes y representantes del paciente.

Anexo 12. Revisiones del informe

Revisores externos (RE)

RE 1

Sección del informe	Comentarios	Respuestas autores
Valoración global	Felicitar al equipo por lo exhaustivo del informe	-
Resúmenes	Excelente y enfoca muy bien el problema	-
Introducción	Lo mismo. Se hace comentario al final en "otros comentarios"	-
Metodología	Lo mismo. Según el comentario final se echa en falta en la página 56 y los términos de búsqueda la palabra "trigger"	-
Resultados	Lo mismo. Se hace comentario al final en "otros comentarios"	-
Discusión	Lo mismo. Se hace comentario al final en "otros comentarios"	-
Conclusiones	Lo mismo. Se hace comentario al final en "otros comentarios"	-
Anexos	Excelente	-
Otros comentarios	El enfoque de la revisión me parece muy pertinente y exhaustiva como se indica previamente. Entiendo si no estoy equivocado que la aproximación al problema es desde una perspectiva de RAM pero en las palabras clave se incluye el aspecto de eventos adversos y la seguridad del paciente por lo que como se indica hay 2 aproximaciones (para mi si no estoy equivocado serían 3): 1. Notificación espontánea 2. Sistemas informáticos mediante algoritmos y modelos IA de HCE sobre textos no estructurados 3. Uso de herramientas tipo Trigger o pistas ej Global trigger tool sobre textos estructurados en la HCE. Este sería otro sistema de recogida como indico y sobre el que hay bibliografía que pongo de ejemplo como la del grupo de MJ Otero. El argumento es que en seguridad de pacientes se requieren varios sistemas de detección de eventos adversos para mejorar la misma.	- El informe se focaliza en el punto 2 comentado por el revisor (también para datos estructurados), y también en el punto 3 (herramientas tipo "trigger") sólo si su detección es automatizada (y verificada manualmente por los expertos) y no realizada de forma exclusivamente manual por los profesionales. Este ha sido el caso de los estudios incluidos de Chan, (2017), Chen (2020) y Quillet (2018). La mayor parte de esta evidencia sobre detección automatizada mediante triggers es anterior a 2013 (fecha límite de inicio de la búsqueda) y se recoge en la RS de Forster (2012), referenciada en la introducción. Posteriormente, la emergencia de los sistemas

	Estos comentarios los dejo a criterio de los autores del documento y la pertinencia o no de tenerlos en cuenta. En cualquier caso, como indico, me parece de nuevo un documento muy exhaustivo y pertinente dado que ahora con la HCE, como aquí en Asturias, es necesario dar un paso más en estos temas. Por supuesto, este tipo de bibliografía debería ser contemplada en los criterios empleados por los autores del documento y la calidad científica que puedan tener. European Journal of Clinical Pharmacology (2023) 79:617–625 https://doi.org/10.1007/s00228-023-03469-5 Evaluation of Global trigger tool as a medication safety tool for adverse drug event detection—a cross-sectional study in a tertiary hospital Ville Valkonen1 • Kaisa Haatainen2,3 • Susanna Saano4 • Miia Tiihonen1 Global Trigger Tool: Proficient Adverse Drug Reaction Autodetection Method in Critical Care Patient Units Amee D Pandya1, Kalan Patel2, Devang Rana3, Sapna D Gupta4, Supriya D Malhotra5, Pankaj Patel6 Indian Journal of Critical Care Medicine (2020): 10.5005/jp-journals-10071-23367 Development and Implementation of an e-Trigger Tool for Adverse Drug Events in a Swiss University Hospital Amina El Saghir 1,*Georgios Dimitriou2,* Miriam Scholer3 Ioanna Istampoulouoglou 1 Patrick Heinrich3 Klaus Baumgartl3 René Schwendimann 4 Stefano Bassetti 2 Anne Leuppi-Taegtmeyer Drug, Healthcare and Patient Safety 2021:13 251–263 European Geriatric Medicine (2022) 13:837–847 https://doi.org/10.1007/s41999-022-00649-x Performance of a trigger tool for detecting adverse drug reactions in patients with polypharmacy acutely admitted to the geriatric ward Nikki M. F. Noorda1 • Bastiaan T. G. M. Sallevelt2 • Wivien L. Langendijk1 • Toine C. G. Egberts2,3 • Eugène P. van Puijenbroek4,5 • Ingeborg Wilting2 • Wilma Knol1 An evaluation of trigger tool method for adverse drug reaction monitoring at a tertiary care teaching hospital Urmila Menat, Chetna K. Desai, Jigar R. Panchal, Asha N. Shah1 Perspect Clin Res 2021;12:33-9.	de IA representa la mayoría de la evidencia identificada.
--	--	--

Hwang, S.-H.; Ah, Y.-M.; Jun, K.-H.; Jung, J.-W.; Kang, M.-G.; Park, H.-K.; Lee, E.-K.; Park, H.-K.; Chung, J.-E.; Kim, S.-H.; et al.
Development and Validation of a Trigger Tool for Identifying Drug- Related Emergency Department Visits.
Int. J. Environ. Res. Public Health 2021, 18, 8572.
<https://doi.org/10.3390/ijerph18168572>

BMC Medical Informatics and Decision Making (2022) 22:146
<https://doi.org/10.1186/s12911-022-01885-8>
Development and assessment of PharmaCheck: an electronic screening tool for the prevention of twenty major adverse drug events
Christian Skalaforis1*, Jean-Luc Reny2, Jérôme Stirnemann2, Olivier Groscurin2, François Eggimann3, Damien Grauser4, Daniel Teixeira4, Megane Jermini1, Christel Bruggmann1, Pascal Bonnabry1,4 and Bertrand Guignard1

Multicenter Test of an Emergency Department Trigger Tool for Detecting Adverse Events
Richard T. Griffey MD, MPH1, Ryan M. Schneider, ACNP1, Brian R. Sharp, MD2, Jeff Pothof, MD2, Marie C. Vrablik, MD, MCR3, Nic Granzella, MD3, Alexandre A. Todorov, PhD4, Lee Adler, DO5
J Patient Saf. 2021 December 01; 17(8): e843–e849.
Doi:10.1097/PTS.0000000000000516.

Otero MJ, Toscano Guzmán MD, Galván-Banqueri M, et al.
Eur J Hosp Pharm 2021;28:e41–e46.
Utility of a trigger tool (TRIGGER-CHRON) to detect adverse events associated with high-alert medications in patients with multimorbidity

J Med Internet Res. 2018 May; 20(5): e198.
doi:10.2196/jmir.9901: 10.2196/jmir.9901
Trigger Tool-Based Automated Adverse Event Detection in Electronic Health Records: Systematic Review
Monitoring Editor: Gunther Eysenbach
Reviewed by Yanshan Wang and Eta Berner Sarah N Musy, MSc, Dietmar Ausserhofer, RN, PhD, René Schwendimann, RN, PhD, Hans Ulrich Rothen, MD, PhD, Marie-Madlen Jeitziener, RN, PhD, Anne WS Rutjes, PhD, and Michael Simon, RN, PhD

RE 2

Sección del informe	Comentarios	Respuestas autores
Valoración global	El estudio aborda un tema muy importante de Salud Pública, ya que Las RAM son una causa frecuente de enfermedad, discapacidad o incluso muerte en nuestro país. El presente estudio revisa los estudios científicos que investigan si estos sistemas realmente son útiles a la hora de detectar Las RAM y reducir sus efectos negativos sobre la salud de las personas, por lo que su objetivo es de gran interés, tanto para las Autoridades Sanitarias, como para los ciudadanos. La inexistencia de estudios de detección de señales en farmacovigilancia, constatan la escasa investigación que ha habido al respecto para mejorar estos sistemas. De ahí que, las conclusiones de este estudio tendrían que ser abordadas, valoradas y evaluadas por las autoridades competentes, elaborando un PLAN DE MEJORA, con la finalidad de incrementar la eficacia de estos sistemas.	-
Resúmenes	Se ajustan a su cometido.	-
Introducción	En la introducción se tratan aspectos fundamentales, como por ejemplo: “La educación y una cultura positiva de notificación de RAM pueden ser muy valiosas para aumentar las tasas de notificación de RAM, por lo que es necesario formar a los equipos de enfermería, medicina, farmacia y otros miembros del equipo sanitario para que asuman la responsabilidad de notificar RAM” Por ello, lo primero que habría que mejorar es la infra notificación, sin datos reales, no se puede mejorar el sistema, ya que los datos que obtenemos son sesgados y carecen de valor. El estudio también expone problemas de salud graves: “El daño a pacientes debido a eventos adversos (EA) de las intervenciones médicas es una causa relevante de muerte y discapacidad a nivel mundial. Anualmente, los hospitales de países de ingresos bajos y medios experimentan una elevada incidencia de eventos adversos (EA) relacionados con una atención sanitaria insegura, alcanzando un total de 134 millones de casos, los cuales contribuyen a 2,6 millones de muertes” “Las RAM son una causa frecuente de enfermedad, discapacidad o incluso muerte [5]. Entre los factores que aumentan el riesgo de que aparezcan se encuentra la polimedicación, la edad de los pacientes, las patologías previas como la insuficiencia hepática y los factores genéticos [6]”.	-

	Parece ser que hasta el momento no se han remediado con el debido ímpetu. Lo que nos lleva a preguntarnos: ¿Por qué no se han estudiado con rigor Las RAM si son realmente un problema de Salud Pública? Es evidente que el estudio de las mismas presenta muchas dificultades técnicas y presupuestarias, entre otras. Además, de poner remedio para solucionarlas, se debería vigilar que las Agencias reguladoras bajo ningún concepto tengan ningún tipo de conflicto de interés con la industria, ya que ésta a su vez se muestra reticente a reconocer cualquier daño, especialmente con las vacunas. Esto puede ocasionar desconfianza ciudadana, cuando observa que no hay completa transparencia. El temor a problemas legales, dificulta también el RECONOCIMIENTO DE LAS RAM y su causalidad con el medicamento en cuestión. No obstante, no se puede obviar que si no se reconoce una RAM nunca se podrá estudiar, ni hacer su seguimiento. De ahí que, para mejorar la eficacia de un medicamento o vacuna sea necesario identificar y reconocer, cuando las haya, Las RAM que puede provocar el medicamento o vacuna en cuestión. Sólo de esta forma la eficacia de un producto y el avance en la ciencia puede mejorar; además de la confianza de los ciudadanos en el mismo.	
Metodología	Se evidencia con la falta de estudios la debilidad de los sistemas de notificación. Dichos estudios serían muy útiles para la prevención, que, sin duda alguna, es un factor de vital importancia para reducir Las RAM. En España, el acceso al cómputo total de reacciones adversas no ha sido posible para el profesional sanitario y los ciudadanos hasta tan sólo hace unos años. Por lo tanto, los profesionales sanitarios no podían conocer el impacto de Las RAM de un medicamento en un momento determinado. En la actualidad existe, pero se debería mejorar la accesibilidad a las notificaciones de Las RAM de un medicamento.	-
Resultados	Es evidente que los sistemas de notificación no funcionan correctamente. Sin embargo, en este estudio, se ofrecen propuestas que se pueden llevar a cabo para mejorarlos. Son ahora las Autoridades Sanitarias competentes, las que se tienen que ponerse manos a la obra.	-
Discusión	Nada que aportar	-
Conclusiones	COMPLETAMENTE DE ACUERDO con las siguientes conclusiones: -Se deberían promover por los organismos correspondientes estudios que evalúen la efectividad y seguridad de los sistemas automatizados de detección de RAM desarrollados, ya que no existen en los últimos 10 años estudios que tengan como objetivo la reducción de RAM y sus consecuencias sobre variables de salud finales. -Se deberían realizar estudios sobre el coste-efectividad de los sistemas automatizados para la detección de RAM. Esto demuestra que este problema no ha sido prioritario para las Autoridades Sanitarias. Si lo fuera, se habría estudiado rigurosamente con anterioridad.	-

	-La implementación de un sistema automatizado de detección de RAM debería llevarse a cabo únicamente tras una meticulosa planificación, en conjunción con los profesionales clínicos. No nos podemos olvidar tampoco en este punto de la importancia del seguimiento y evaluación de las RAMs.	
Anexos	Adecuados	-
Otros comentarios	VALORACIÓN GENERAL Hay que evaluar y mejorar los sistemas de notificación para que sean útiles a la hora de detectar las RAM y reducir sus efectos negativos sobre la salud de las personas. Es evidente, que hoy en día, no son útiles. Por ello, creemos que habría que diseñar un Plan de Mejora, que incluyesen algunas de las conclusiones del presente estudio y otras que sugerimos. PLAN DE MEJORA PREVENCIÓN • Revisar la historia Clínica antes de la prescripción de un medicamento. • Reconocer una RAM. La resistencia a identificarla, sobre todo cuando se producen tras la administración de una vacuna, dificulta su investigación y deja al ciudadano completamente desamparado. • Para facilitar la notificación de RAM se debe mejorar la formación del profesional sanitario: conocimiento y concienciación de notificación, ya que son los que normalmente notifican y deben saber cómo proceder. Además, si reciben una buena formación serán capaces de detectar errores del sistema y aportar posibles soluciones. • Contar con suficiente personal sanitario para evitar errores de medicación provocados por una sobrecarga laboral. • Intensificar las RAM (20-30%) que son potencialmente prevenibles, intensificando estrategias de salud. • Informar a los ciudadanos de la existencia de un sistema de notificación, ya que generalmente desconocen que existe, y/o no conocen su funcionamiento. MAYOR INVERSIÓN Mayor inversión en Sistemas informáticos que hagan seguimiento de las RAM y evolución de los pacientes. Generalmente cuando se notifica no se hace un seguimiento de las mismas. Además, utilizar la IA supervisada por técnicos supondría un gran avance técnico tanto para la recopilación de los datos, como su interpretación. EFICACIA DE LA EVALUACIÓN	-

	Cuando se notifica una sospecha, la evaluación debe de ser eficaz. Las Agencias Reguladoras deben de actuar con total transparencia. Una vez evaluada la sospecha, se debe informar a los profesionales sanitarios, que son los que están recetando los medicamentos y notificando sus posibles RAMs. Ante cualquier sospecha, se debería interrumpir el medicamento o vacuna administrada, para evitar posibles males mayores. Además, sería muy beneficioso hacer extensible a todo el territorio español el proyecto de innova- Saúde, desarrollado por el Servicio Gallego de Salud u otras iniciativas similares, para identificar mejor la información y generar alertas inteligentes. SEGUIMIENTO médico prolongado en el tiempo de las RAM. REPARACIÓN Fondo de Compensación. La preocupación por posibles consecuencias legales, se mitigaría con la existencia de un fondo de compensación de daños por medicamentos. El paciente en ningún caso debe de asumir el riesgo, y si lo asume, el sistema debe repararlo. Es primordial por ello, ponerse a trabajar para mejorar la infranotificación de Las RAM y su seguimiento, así como diseñar un Plan de Mejora, ya que el daño a pacientes debido a eventos adversos (EA) de las intervenciones médicas es una causa relevante de muerte y discapacidad a nivel mundial. Por lo tanto, debe de ser un tema prioritario de Salud Pública.	
--	---	--

RE 3

Sección del informe	Comentarios	Respuestas autores
Valoración global	Se trata de un informe de evaluación de tecnologías sanitarias bastante completo que responde a una pregunta de investigación adecuada y pertinente en el momento actual	-
Resúmenes	Nada que añadir	-
Introducción	Cubre todos los aspectos relevantes Nada que añadir	-

Metodología	La metodología empleada es la adecuada para responder a la pregunta de investigación	-
Resultados	Página 82: Los/as profesionales de farmacia deben comprobar los errores e introducir manualmente el resultado en la HCE. Esto puede ser una limitación de los sistemas automáticos actuales de detección de RAM. ¿Esta frase de la limitación de los sistemas automáticos al tener que comprobar los errores e introducirlos manualmente en la HCE se basa en algún estudio? O son percepciones? La duda me surge al no estar referenciada esta afirmación en el texto. Pag 85: probabilidades de reportar RAM si el proceso fuera mejorado mediante la adopción de una herramienta electrónica automatizada y si la notificación fuera obligatoria [67] En España la notificación de RAM es obligatoria otra cosa es la falta de tiempo y de formación en relación a la importancia de notificar las RAM de los profesionales sanitarios Se recomienda hacer esta aclaración en el texto. Pag 88: Al compararlos, el modelo de no consentimiento tiene más probabilidades de explotar plenamente la información sanitaria para detectar las RAM Entiendo que son los modelos de consentimiento los que tienen más probabilidades de explotar plenamente la información	- Se ha eliminado la frase señalada en la página 82 y modificado el texto en las 85 y 88 en el sentido señalado por el revisor.
Discusión	Nada que añadir	-
Conclusiones	Me parecen coherentes con los resultados obtenidos del informe, lamentablemente no es posible formular recomendaciones a partir de las conclusiones del informe como indican los autores	-
Anexos	Correcto	-
Otros comentarios	Nada que añadir	-

Revisiones de representantes sociedades científicas (RSC)

RSC 1

Sección del informe	Comentarios	Respuestas autores
Valoración global	Globalmente considero que el informe es completo, bien justificado, y que se han revisado todas las referencias relevantes para describir y analizar el estado del arte de los sistemas de detección automática de RAMs. En referencia al título, tengo un comentario editorial menor. En la traducción al inglés "Título Effectiveness, safety and cost-effectiveness of automated mated systems for detecting adverse drug reactions" sobra la palabra "mated".	Se ha corregido el error en la traducción del título
Resúmenes	Los resúmenes en lenguaje no especializado están bien redactados y se comprenden bien. El resumen ejecutivo contiene los puntos y conclusiones más importantes. En cuanto al resumen en lenguaje no especializado, cuando se indica: "Cuando un/a profesional sanitario/a sospecha que un/a paciente puede estar padeciendo una RAM, puede rellenar un formulario describiendo el caso, y esta información es revisada por expertos/as en farmacia y farmacología, junto al resto de datos clínicos del/la paciente, para determinar si se trata de una RAM real. Esta información queda almacenada en última instancia en la base de datos FEDRA (Farmacovigilancia Española de Datos de Reacciones Adversas)", se podría añadir que esta información es evaluada por expertos en farmacia y farmacología clínica, y que en última instancia es reportada/informada a las autoridades regulatorias nacionales en materia de medicamentos, quienes evalúan esta información, la almacenan en última instancia en la base de datos FEDRA, y monitorizan de forma continua la seguridad de los medicamentos analizando si es preciso tomar medidas. Es importante que se destaque el papel de las Autoridades Regulatorias, como la AEMPS, en la vigilancia de la seguridad, sobre todo de cara a las personas no profesionales que puedan leer este resumen.	Se ha modificado la frase para incluir la sugerencia del revisor.
Introducción	La introducción es clara y está muy bien justificada. No tengo comentarios adicionales a este apartado.	-
Metodología	La metodología es completa, correcta y precisa. Aporto algunos comentarios: Se realizó una búsqueda de la literatura en las bases de datos MEDLINE, EMBASE, COCHRANE y Web of Science Core Collection hasta junio de 2023. Propongo añadir justificación de por qué se utilizaron estas bases de datos, y no otras (por ejemplo, en término de áreas y publicaciones cubiertas con dichas bases de datos).	- Se armonizado la descripción del periodo de búsqueda. - En la mayoría de estudios no se verifica la relación de causalidad entre los eventos adversos detectados (EAM) y la toma de medicación, es decir, que realmente fueran

	Por otra parte, en el protocolo no se determinó una fecha inicial para la inclusión de estudios, pero dado el rápido desarrollo de las tecnologías evaluadas se decidió incluir estudios publicados después de 2013. No obstante, en el resumen de específica que se excluyeron las publicaciones de antes de 2014. Sugiero armonizar la forma de describir el periodo incluido. En el texto se habla indistintamente de EAM/EA/RAM. Sugiero armonizar la terminología, y en caso de requerir utilizar distintos términos, explicar en cada caso a qué hace referencia cada término y qué se medía con dicha variable, a efectos de claridad. Por ejemplo, a qué nos referimos cuando hablamos de medir EAM, o cuando medimos RAM, y por qué el estudio utiliza uno u otro como variable de medida.	RAM, por lo que se ha mantenido el término usado en cada estudio.
Resultados	Los resultados son claros y están descritos de manera amplia y completa. Sugiero algunos comentarios: Apartado IV.1.4. Efectividad y seguridad: Se utilizan varias siglas, que si bien se han explicado antes de este apartado, o en los anexos, igual merecería la pena indicar a qué hacen referencia la primera vez que se utilizan, a efectos de facilitar la lectura de los resultados. Uno de los estudios encontrados de efectividad y seguridad se realizó con cuestionarios que debían rellenar los pacientes. Los resultados reportados por los pacientes son especialmente relevantes, especialmente en farmacovigilancia. Por ello, se podría ampliar un poco más la metodología de este estudio, indicando si se trató de un cuestionario diseñado para el estudio, o era un cuestionario asistencial, y en qué variables mostró rendimiento el PLN utilizado. Comentario editorial menor: Página 86: Cuando se indica "En los estudios incluidos, se han identificado varios problemas relacionados con el uso de HCE para detectar RAM. Estos problemas incluyen la falta de estandarización entre sistemas de HCE, la especificidad en las opciones de ingreso de datos, la documentación incompleta e inexacta, y la fatiga por alerta", sugeriría añadir referencia(s).	- Se han explicado las abreviaturas referidas a los parámetros estadísticos, no así las conceptuales explicadas previamente en el texto (RAM, EAM, PLN, etc). - Se ha añadido que el cuestionario online usado por Iivainen (2021) fue desarrollado en un estudio previo específicamente para esos pacientes. Se explicita que, dado el pequeño tamaño muestral solo pudo evaluarse la presencia de cualquier RAM y no de síntomas específicos. - Se han añadido referencias en el comentario de la página 86.
Discusión	Discusión, página 95: "V.1 Efectividad y seguridad": Cuando se discute el único estudio de efectividad encontrado, haría más énfasis en la importancia de hacer caso a la alerta y de interpretarla bien. Considero que este punto es una de las conclusiones más importantes que se derivan de este estudio. En el momento del alta, la tasa de discontinuación sí fue mayor en el grupo de intervención, pero solo cuando se analizaron únicamente los médicos que respondieron al cuestionario post-alarma. Los sistemas de detección automatizada de RAM pueden ayudar a los profesionales a incrementar la tasa de detección de RAM. Pero como bien se explica en el informe, dichas RAMs deben ser validadas por profesionales, y lo que es más importante, debe redundar en una modificación de la actuación del profesional, lo que pasa por establecer una alerta que pueda informar antes de la	- Se ha añadido una frase al final del primer párrafo para señalar la importancia de una adecuada atención e interpretación de las alarmas. - Se ha mejorado la redacción del comentario señalado en la página 99.

	prescripción. Resulta esencial, y lo muestra de forma preliminar este estudio, que los profesionales vean las alertas y las interpreten correctamente. Comentario editorial menor: Página 99: Esta frase no se entiende bien: "Además de ello, para minimizar en la medida de lo posible es la existencia de sesgos en los conjuntos de datos utilizados para entrenar los algoritmos de sistemas de IA en salud, ya que estas bases de datos pueden reforzar los sesgos existentes"	
Conclusiones	Como se indica en la página 93, dada la ausencia de estudios que evaluaran variables finales de efectividad y seguridad, las necesidades de investigación están completamente por cubrir. Merecería la pena destacar este punto en las conclusiones, reforzando la idea de que se debería invertir más fondos en la promoción de la investigación dentro de esta área, en la generación de evidencias, y en el desarrollo de un sistema fiable que pueda ser interoperable en todo el SNS. Este punto es imprescindible antes de que pueda implantarse un sistema de automatizado de detección de RAMs en práctica asistencial. Así mismo, también debería destacarse que, en el momento actual, y con el estado del arte que bien se ha revisado en este trabajo, no puede implantarse un sistema generalizado de detección automática de RAMs en práctica asistencial.	-
Anexos	Los anexos se entienden bien, son correctos, y no tengo comentarios adicionales.	-
Otros comentarios	No tengo comentarios adicionales.	-

RSC 2

Sección del informe	Comentarios	Respuestas autores
Valoración global	Muy buena.	-
Resúmenes	Los considero ambos adecuados y el ejecutivo bien estructurado.	-
Introducción	Valorar incorporar la experiencia del módulo de seguridad clínica de la CCAA Cataluña en las experiencias nombradas en el punto I.3 (página 38) disponible en https://medicaments.gencat.cat/ca/professionals/seguretat/seguretat-clinica-recepta-electronica/ O la del Servicio Navarro de Salud- Osasunbidea (SNS-O) http://www.navarra.es/home_es/Actualidad/Sala+de+prensa/Noticias/2018/01/02/Pr+emiada+herramienta+informatica+del+SNS-O+que+mejora+seguridad+farmacoterapia.htm	- Se han incluido las iniciativas en Cataluña y Navarra.

Metodología	Ninguna aportación.	-
Resultados	Ninguna aportación.	-
Discusión	Ninguna aportación.	-
Conclusiones	Ninguna aportación.	-
Anexos	Ninguna aportación.	-
Otros comentarios	Ninguna aportación.	-

RSC 3

Sección del informe	Comentarios	Respuestas autores
Valoración global	El informe cumple adecuadamente su cometido y supone una puesta al día de gran interés y utilidad para todos los actores implicados.	-
Resúmenes	Me parecen adecuados y comprensibles.	-
Introducción	Para justificar la alta prevalencia de polimedicación en población española, se sugiere este estudio del grupo de trabajo de utilización de fármacos de la semFYC con la base de datos BIFAP: Evolution of polypharmacy in a spanish population (2005-2015): A database study - PubMed (nih.gov)	-
Metodología	Me parece muy adecuada.	-
Resultados	Son coherentes con la estrategia de búsqueda y hallazgos de la revisión sistemática.	-
Discusión	Resume muy bien el estado del conocimiento y los retos a resolver en el futuro.	-
Conclusiones	Me parece adecuadas y coherentes con los resultados.	-
Anexos	Sin comentarios.	-
Otros comentarios	Sin comentarios.	-

RSC 4

Sección del informe	Comentarios	Respuestas autores
Valoración global	En conjunto creo que es una excelente revisión que refleja claramente el estado actual de la evidencia existente en cuanto a los sistemas automatizados de detección de reacciones adversas a medicamentos. Muestra de forma precisa la importancia del	-

	tema, la escasa evidencia existente, el posible impacto futuro y marca las líneas para futuros desarrollos.	
Resúmenes	Explica de forma clara y inteligible para un público general y no especializado los objetivos, la metodología utilizada y los resultados encontrados	-
Introducción	Refleja la magnitud del problema y la necesidad existente de mejorar los circuitos actuales de detección de RAM. Explica el potencial de IA como herramienta de ayuda para mejorar la detección de las RAM. En el apartado regulatorio creo que sería importante actualizarlo señalando el acuerdo político alcanzado entre el Parlamento Europeo y el Consejo sobre la Ley de Inteligencia Artificial (Ley de IA) en Diciembre del 2023. Constituye el primer marco jurídico global en materia de inteligencia artificial en todo el mundo. Marca las bases para identificar posibles riesgos identificables y para fomentar la innovación responsable en este campo. En esta ley queda definido que los sistemas de IA implicados en la toma de decisiones en salud son de alto riesgo. deberán cumplir requisitos estrictos, tales como sistemas de mitigación de riesgos, conjuntos de datos de alta calidad, registro de actividad, documentación detallada, información clara a los usuarios supervisión humana y un alto grado de solidez, precisión y ciberseguridad. Los sistemas de IA en salud serán considerados dispositivos sanitarios y como tales sometidos a la clasificación y control regulatorio de cualquier otro dispositivo sanitario. Cada algoritmo de IA deberá contar con una ficha técnica que informe sobre las características técnicas, intención de uso (primario-secundario), contraindicaciones, consideraciones éticas, datos de población de entrenamiento y variables usadas, datos de evaluación (fecha, poblaciones de evaluación externas, métricas, resultados, limitaciones) y criterios de monitorización (mecanismos, fechas, puntos de mejora detectados) https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_es https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2022)729512	- Se ha incluido la mención a la Ley de IA europea.
Metodología	La metodología me parece que está bien estructurada y explicada. En la evaluación crítica de la parte de sesgos creo que puede ser importante destacar que los algoritmos de IA aprenden a tomar decisiones en función de los datos de entrenamiento de decisiones humanas y estas pueden ser sesgadas o reflejar desigualdades históricas o sociales, incluso si se eliminan variables sensibles como el género, la raza o la orientación sexual. Es importante incluir expertos en ética y desigualdades en los equipos interdisciplinarios encargados de diseñar, validar y trasladar a la clínica los algoritmos de IA.	- En metodología se describe la evaluación del riesgo de sesgo de los estudios incluidos, el posible sesgo introducido en el aprendizaje de los modelos de IA se comenta en el apartado de aspectos éticos. - Respecto al comentario sobre el estudio económico, estamos de acuerdo con esta apreciación quedando reflejada en la Discusión del informe.

	- Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. N Engl J Med. 2020 Aug 27;383(9):874-882. doi: 10.1056/NEJMms2004740. Epub 2020 Jun 17. PMID: 32853499. - Buslón N, Cortés A, Catuara-Solarz S, Cirillo D, Rementeria MJ. Raising awareness of sex and gender bias in artificial intelligence and health. Front Glob Womens Health. 2023 Sep 6;4:970312. doi: 10.3389/fgwh.2023.970312. PMID: 37746321; PMCID: PMC10512182. Como sugerencia creo que en el estudio económico, ante la ausencia de estudios de coste-efectividad, y al centrarse en estudio de costes directos quizá sea oportuno incluir la valoración de impacto de varias empresas y no una sola. Este permitiría dar una horquilla aproximada. Ante la ausencia de varias fuentes quizá sea más correcto no hacer referencia a los costes directos.	
Resultados	En base a lo expuesto anteriormente creo que sería aconsejable evitar la referencia a una única propuesta de valoración económica en base a un único presupuesto. Creo que sería mejor para futuras revisiones incluir una horquilla de precios en base a la consulta de diferentes empresas especializadas en el tema.	- Se ha comentado en la discusión.
Discusión	En la discusión creo que es importante enfatizar el papel fundamental de la implicación y formación transdisciplinaria en la detección de RAM, destacando a todos los profesionales sanitarios de los diferentes ámbitos y niveles asistenciales. Es igualmente importante incluir formación en la prevención, detección y abordaje de las RAMSs en el ámbito de los grados, post-gradados y formación sanitaria especializada. Creo que también se puede enfatizar la implicación de los pacientes mediante el uso de PREMS-PROMS, resultados y experiencias reportados por los pacientes, en la detección de posibles RAMs y como medidas básicas en a la valoración de los resultados en salud. Estos datos deben poder diferenciarse claramente de los reportados por los profesionales asistenciales. Zoni Matta AC, Pichiule Castañeda M, Neira León M, Hoyos Miller J, Illarramendi P, Ballester Santiago M, Suñol Sala R, van den Berg M, Aguilera Guzmán M. International Patient-Reported Indicators Surveys: The PaRIS study protocol in Spain [Encuesta Internacional de Indicadores de Salud Referidos por los Pacientes: el protocolo del Estudio PaRIS en España.]. Rev Esp Salud Publica. 2023 Sep 7;97:e202309072. Spanish. PMID: 37970989; PMCID: PMC10540904. Como posibles líneas futuras de investigación señalaría la creación de grandes bases de datos nacionales e internacionales que permitan la detección de "patrones de	- Se ha incorporado a la discusión el comentario del revisor sobre la formación transdisciplinar y universitaria.

	conducta” asociados a alto riesgo de RAM y el desarrollo posterior de algoritmos de IA que sirvan de cribado poblacional y alerta precoz de alto riesgo de RA: Estas alertas deben estar incluidas en el flujo de trabajo clínico-asistencial-farmacéutico y para evitar la fatiga por exceso de alertas limitadas a la de mayor impacto clínico y a los periodos críticos. Otro punto interesante es que se está empezando a hablar del concepto de IA-verde y sostenible y que busca dar respuesta e integrar los criterios de responsabilidad medioambiental y sostenibilidad en los sistemas de inteligencia artificial. Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. WIREs Data Mining and Knowledge Discovery, 13(4), e1507. https://doi.org/10.1002/widm.1507	
Conclusiones	Me parecen adecuadas y correctas. Creo que sería recomendable un periodo de actualización máximo de 2 años ya que los cambios esperables en este campo van a ser destacables a muy corto plazo.	-
Anexos	Ningún comentario a añadir	-
Otros comentarios	- Basile AO, Yahi A, Tatonetti NP. Artificial Intelligence for Drug Toxicity and Safety. Trends Pharmacol Sci. 2019 Sep;40(9):624-635. doi: 10.1016/j.tips.2019.07.005. Epub 2019 Aug 2. PMID: 31383376; PMCID: PMC6710127. - Martin GL, Jouganous J, Savidan R, Bellec A, Goehrs C, Benkebil M, Miremont G, Micallef J, Salvo F, Pariente A, Létinier L; French Network of Pharmacovigilance Centres. Validation of Artificial Intelligence to Support the Automatic Coding of Patient Adverse Drug Reaction Reports, Using Nationwide Pharmacovigilance Data. Drug Saf. 2022 May;45(5):535-548. doi: 10.1007/s40264-022-01153-8. Epub 2022 May 17. PMID: 35579816; PMCID: PMC9112264. CopyDownload .nbib Fo - Abdallah Abou Hajal & Ahmad Z. Al Meslamani (2023) Overcoming barriers to machine learning applications in toxicity prediction, Expert Opinion on Drug Metabolism & Toxicology, DOI: 10.1080/17425255.2023.2294939 - Sarkar C, Das B, Rawat VS, Wahlang JB, Nongpiur A, Tiewsoh I, Lyngdoh NM, Das D, Bidarolli M, Sony HT. Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. Int J Mol Sci. 2023 Jan 19;24(3):2026. doi: 10.3390/ijms24032026. PMID: 36768346; PMCID: PMC9916967.	-

RSC 5

Sección del informe	Comentarios	Respuestas autores
Valoración global	Responde adecuadamente al interés despertado por la irrupción de la inteligencia artificial a través del aprendizaje automático que utiliza algoritmos y aprendizajes previos para realizar predicciones aplicado a la farmacovigilancia de productos ya comercializados y fármacos en desarrollo.	-
Resúmenes	Son comprensibles, completos y explican con claridad lo esencial del Informe.	-
Introducción	Es completa y permite un acercamiento al tema abordado. Desde el punto de vista de la Psiquiatría y Salud Mental, quizá hubiera sido interesante aportar alguna información específica como también se podría hacer con especificidades de otras Especialidades . Concretamente referir que los factores psicológicos pueden influir en la probabilidad de que un paciente presente reacciones adversas a un fármaco. Los factores psicológicos, como el estrés, la ansiedad, la depresión y otros trastornos mentales, pueden afectar la respuesta del cuerpo a los medicamentos. Por ejemplo, el estrés crónico puede alterar el equilibrio hormonal y la función inmunológica, lo que puede aumentar la sensibilidad a los efectos secundarios de los medicamentos. Además, los trastornos de ansiedad y depresión pueden afectar la percepción del dolor y la respuesta a los medicamentos analgésicos, lo que puede aumentar el riesgo de reacciones adversas. Además, los factores psicológicos, como las expectativas y creencias sobre los medicamentos, pueden influir en la forma en que un paciente percibe y reporta los efectos secundarios. Por ejemplo, un paciente que tiene una fuerte creencia de que un medicamento causará efectos secundarios negativos puede ser más propenso a experimentarlos, incluso si el medicamento en sí no tiene propiedades que los causen (efecto nocebo). Medicamentos y monitorear de cerca cualquier cambio en la respuesta del paciente. Aquí hay algunos estudios y publicaciones al respecto: 1. Tylee, A., & Gandhi, P. (2005). The importance of somatic symptoms in depression in primary care. Primary care companion to the Journal of clinical psychiatry, 7(4), 167-176. 2. Barsky, A. J., Saintfort, R., Rogers, M. P., & Borus, J. F. (2002). Nonspecific medication side effects and the nocebo phenomenon. Jama, 287(5), 622-627.	- Se ha incluido un comentario y referencias sobre la influencia de las variables psicológicas en la ocurrencia de RAM.

	3. Rief, W., & Avorn, J. (2006). Persistent medical symptoms without objective evidence: a systematic review of patients' expectations, placebo effects, and somatic attributions. <i>Harvard review of psychiatry</i>, 14(2), 80-90. 4. Benedetti, F. (2009). Placebo effects: from the neurobiological paradigm to translational implications. <i>Neuron</i>, 84(3), 623-637. 5. Colloca, L., & Miller, F. G. (2011). The nocebo effect and its relevance for clinical practice. <i>Psychosomatic medicine</i>, 73(7), 598-603.	
Metodología	La metodología utilizada para evaluar los aspectos éticos, legales, organizativos, sociales y ambientales de un sistema de interceptación del proceso de prescripción y soporte a la prescripción farmacéutica parece ser exhaustiva y rigurosa. El hecho de haber seguido un algoritmo para evaluar estos aspectos demuestra un enfoque sistemático y estructurado. Además, se destaca que se realizaron búsquedas manuales y consultas con las partes interesadas, lo que indica un esfuerzo por recopilar diferentes perspectivas y opiniones relevantes. La inclusión de revisiones sistemáticas, revisiones de alcance, estudios cualitativos, estudios observacionales y estudios de métodos mixtos en la revisión de la literatura también muestra una amplia cobertura de fuentes de información.	-
Resultados	En cuanto a los resultados presentados, se destaca la falta de información necesaria para llevar a cabo una evaluación económica completa. Esto limita la comprensión de los costos y beneficios asociados con la implementación y mantenimiento del sistema automatizado de detección de RAM. Sería útil contar con datos más detallados y completos para tener una visión más clara de la viabilidad económica de esta intervención aunque ya se explica que no los hay. En relación a los aspectos éticos y legales, se menciona la importancia de la privacidad en los sistemas automatizados de detección de RAM, así como la necesidad de desarrollar herramientas de protección de la privacidad y abordar las amenazas de ciberseguridad. Sin embargo, no se proporcionan recomendaciones concretas o medidas específicas para abordar estos desafíos éticos y legales aunque no sea objetivo de este informe En términos de aspectos organizativos, se señala que los sistemas tecnológicos de notificación de RAM integrados en el flujo de trabajo pueden facilitar la notificación de RAM. Sin embargo, también se mencionan obstáculos como la falta de	- Respecto al comentario sobre los datos necesarios para el estudio de la viabilidad económica, efectivamente, no se dispuso más que de los datos presentados en el presente informe, abordándose este aspecto en el apartado de Discusión. - Se ha incluido una sección dentro del apartado de resultados de los Aspectos éticos, legales, organizativos y sociales que sintetiza las sugerencias o propuestas de mejora que deben tenerse en cuenta al implantar estos sistemas.

	estandarización entre sistemas de HCE, la documentación incompleta e inexacta, y la fatiga por alerta. Sería útil contar con recomendaciones más concretas sobre cómo abordar estos desafíos organizativos y mejorar la integración de los sistemas automatizados de notificación de RAM en los sistemas de gestión de medicamentos y registros electrónicos.	
Discusión	Es relevante poner de manifiesto que la PLN aplicada a la narrativa de las HCE, no es, a día de hoy, una opción real pues se encuentra en un nivel de desarrollo muy inicial, y que sin embargo cuando la fuente de información para la IA son los datos estructurados entonces si se detecta eficazmente la RAM.	-
Conclusiones	Son Conclusiones derivadas del trabajo anterior que enfrían las expectativas que pudieran tenerse sobre la cercana aplicación de la automatización de la detección de RAM a través de la IA. En primer lugar, se menciona que no se identificaron estudios que evaluaran la efectividad y seguridad de los sistemas automatizados de detección de RAM en la reducción de RAM y sus consecuencias sobre variables de salud finales, excepto un único estudio que evaluó la implementación hospitalaria de un sistema de procesamiento de lenguaje natural (PLN). Esta falta de estudios limita la capacidad de sacar conclusiones sólidas sobre la efectividad y seguridad de estos sistemas y resalta la necesidad de más investigación en este campo. En cuanto al rendimiento diagnóstico de los métodos de inteligencia artificial (IA) aplicados a los datos de historias clínicas electrónicas (HCE), se menciona que en general muestran un buen rendimiento global, pero que se centran principalmente en la detección de medicación y no en la detección específica de RAM en los textos clínicos. Además, se destaca la variabilidad en el rendimiento de los modelos cuando se estudian RAM específicas. Esto resalta la necesidad de mejorar la capacidad de detección de RAM en los textos clínicos y realizar validaciones externas para verificar la ocurrencia de RAM reales. En relación a la implementación de un sistema automatizado de detección de RAM, se menciona la importancia de una meticulosa planificación en conjunto con los profesionales clínicos y considerando aspectos como el área clínica donde se maximice su efectividad y eficiencia, las expectativas y preocupaciones de los profesionales y el impacto en su flujo de trabajo. Sin embargo, no se brindan recomendaciones específicas sobre cómo abordar estos aspectos y lograr una implementación exitosa. En cuanto al coste-efectividad de los sistemas automatizados de detección de RAM, se menciona que no se identificó evidencia al respecto. Esto es una limitación	-

	importante, ya que la evaluación del coste-efectividad es crucial para tomar decisiones informadas sobre la implementación de estos sistemas.	
	En general, el capítulo de conclusiones proporciona una visión general de los hallazgos y recomendaciones basadas en la revisión realizada.	
Anexos	Completan muy bien los pasos y estrategias explicadas previamente en el Informe.	-
Otros comentarios		-

RSC 6

Sección del informe	Comentarios	Respuestas autores
Valoración global	La valoración global de la revisión es muy buena. Se trata de un tema de actualidad, conocido por bastantes especialistas de nuestra especialidad y del que se asume un alto rendimiento (impulsado por la propaganda de las empresas que venden algunos de los sistemas), pero no se dispone de evidencias contrastadas. La metodología es correcta y las conclusiones responden a los objetivos. En general, constituye una buena ayuda a la toma de decisión acerca de la implementación o de la incorporación a la cartera de servicios del Sistema Público de Salud.	-
Resúmenes	•En las Siglas y acrónimos hay alguna errata (ver documento en pdf adjunto) y se repiten algunos (ver documento adjunto). •El Resumen en lenguaje no especializado es demasiado técnico. Creo que sería de utilidad revisarlo “a la baja” y pasarle alguna herramienta de legibilidad tipo Flesh-Kincaid antes de dar la redacción definitiva. •En el Resumen ejecutivo se habla de “imposibilidad de revisar manualmente”, y en realidad la expresión debe ser “dificultad de...” •En el apartado de Método correspondiente al Resumen debe hacerse constar (referido a las dos revisiones sistemáticas) si se restringió por idioma, si se hizo seguimiento de referencias, si se analizó el sesgo de publicación (o si se pretendía hacerlo), mencionar el análisis de riesgo de sesgo (o si se pretendía hacerlo). •La redacción del apartado de Método debe estar en pasado; revisar. •Cuando se habla de la RS de coste-efectividad no se menciona en qué bases se hizo la búsqueda; cabe pensar que se trata de las mismas mencionadas en el apartado anterior, pero debe señalarse. •En el apartado de estudio de costes se habla de “prescripción farmacéutica” y parece quererse decir “prescripción de fármacos”. Aclarar.	- Se han corregido/modificado los errores o inexactitudes mencionados por el revisor. - Respecto a la RS de coste-efectividad, se menciona que se siguió la misma estrategia de búsqueda que para la RS de efectividad y seguridad. Tratándose de un resumen consideramos que es información suficiente. - Se ha calculado e incluido el NNT en los resultados del estudio de efectividad de Smith (2018).

	•En el apartado de Resultados del Resumen, cuando se habla de efectividad y seguridad se dan los porcentajes de reducción de la tasa de prescripción y algún otro. ¿Podría calcularse un NNT? Pienso que sería de cierta utilidad para el lector. •En el apartado de Coste-efectividad, seguramente es mejor hablar de que “no se encontraron estudios” que de “no se identificó evidencia”. •En el epígrafe de Aspectos éticos, legales, no queda muy claro si las cuestiones planteadas se desprenden de la revisión o son conclusiones. Si es así, deben cambiar de apartado.	
Introducción	•En el punto I.1 se habla de “...intervenciones médicas...” y sería mejor hablar de “...intervenciones sanitarias...” ya que no son siempre exclusivamente médicas. •Se habla de la incidencia de EA en los países de ingresos bajos y medios pero de los países de ingresos altos. Los datos proporcionados provienen de un estudio de la OMS donde también se dan los datos de daños en países de ingresos elevados (https://apps.who.int/gb/ebwha/pdf_files/EB144/B144_29-en.pdf - Global Patient Safety Action Plan 2021–2030 Towards Eliminating Avoidable Harm in Health Care . Geneva: World Health Organization; 2021). •En los ejemplos de RAM tipo B debe precisarse que las erupciones cutáneas son con “algunos” antibióticos. •Cuando se habla de la revisión hecha en España a través del CMBD se menciona a los “esteroides de la corteza suprarrenal” cuando es más correcto hablar de corticoides o corticosteroides. •Un poco más adelante, al mencionar el informe anual de la AEMPS se habla de medicamentos “para trastornos generales”, sin que se sepa muy bien a qué grupo terapéutico se refiere. Tal vez merecería la pena identificarlos mejor con la clasificación ATC (N para sistema nervioso, M para sistema musculoesquelético y <i>¿¿</i> para los “trastornos generales”). •En el 2º párrafo de la pág. 35 se dice “...posibilidad de realizar una asociación causal...” cuando lo correcto sería hablar de “aproximarse a una relación causal”. La inferencia de causalidad mediante este sistema de notificación voluntaria es muy dudosa. •Al final de este mismo párrafo se dice que imposibilidad de calcular la incidencia de RAM tiene que ver con el desconocimiento del denominador cuando en realidad lo que se desconoce es el numerador; el denominador puede venir dado con una buena aproximación por el nº de unidades consumidas/ventas, mientras que el numerador tiene el grave hándicap de que únicamente se conoce lo comunicado, no el total de RAM producidas. •Al final del tercer párrafo de la pág. 35 se menciona que se pueden procesar grandes cantidades de datos de forma “eficiente” y sería mejor hablar de forma “efectiva”, ya que de momento no hay cuestiones económicas de por medio.	- Se han corregido/modificado los errores o inexactitudes mencionados por el revisor, aunque se ha mantenido el formato de las preguntas de investigación.

	•Al final de la pág. 36 se dice que el objetivo es que las máquinas “comprendan”. Creo que no es el término adecuado; tal vez sería mejor “procesaran” o algo similar. •En el primer párrafo de la pág. 37 debe mejorarse la sintaxis; aparecen sucesiones de dos puntos (en los que uno incluye a los otros), mayúsculas detrás de punto y coma, etc. Mejorar la sintaxis del párrafo. •En el primer párrafo de la pág. 38 se usa como abreviatura e.g. Lo correcto en español sería p.e. •En el apartado II.3 Preguntas de investigación se formula ésta con formato PICO pero no se define cómo medir la efectividad, aspecto que forma parte del formato PICO. Tal vez debería mejorarse la definición de los outcomes de resultados.	
Metodología	•En el apartado III.1.3 debe mencionarse si se hizo seguimiento de referencias. •En el apartado III.4.1 p. 51, debe mejorarse la redacción del 2º párrafo, no queda nada claro para el lector común y probablemente tampoco para el especializado. •En el apartado III.8.2 se menciona como participación de sociedades científicas a FEFE (Federación Empresarial de Farmacéuticos Españoles). No me parece que se trate de una sociedad científica ni que entre sus fines se encuentre la mejora de la profesión u objetivos científicos. La eliminaría de este apartado, colocándola con la Industria o creando una categoría específica. •En el mismo apartado se cita a AESEG (Asociación Española de Medicamentos Genéricos) y tampoco creo que se halle entre las sociedades científicas, más bien debe situarse en la participación de la Industria. •Debe mencionarse si se pretendía hacer análisis de riesgo de sesgo de publicación y de qué forma.	- Se ha modificado el párrafo mencionado de la página 51. - Las dos asociaciones mencionadas se han incluido en el apartado de industria. - Se ha comentado que, dado que no se realizó MA de resultados, no se analizó el sesgo de publicación.
Resultados	Nada que mencionar, excepción hecha de que debe decirse el resultado del sesgo de publicación: si se hizo y, en caso contrario, la causa de que no se hiciera.	-
Discusión	Nada que mencionar.	-
Conclusiones	Nada que mencionar.	-
Anexos	•En los representantes de sociedades científicas aparezco como revisor del protocolo y de la versión final del informe. No recuerdo haber sido consultado para la revisión del protocolo en la reunión que tuvo lugar en abril de 2023. Debe corregirse.	
Otros comentarios	•En todo el texto debe sustituirse el punto decimal por la coma, más natural en nuestro idioma. Además, la separación en los millares debe hacerse con un punto (125.000) y no con un espacio (125 000). •En la Bibliografía debe utilizarse las abreviaturas estándar de las revistas. P. e. Clin Pharmacol Ther y no Clin. Pharmacol. Ther. En los Anexos están correctamente citadas.	Aunque la separación de decimales con comas y de millares con puntos es más habitual en España, también está reconocido en la lengua española el uso de la separación de decimales con puntos y de millares con espacios finos. La ventaja del uso de este último tipo de ortografía de las expresiones numéricas, reconocida internacionalmente, es

	que facilita la posterior difusión de los resultados en ámbitos internacionales.
--	--

RSC_7

Sección del informe	Comentarios	Respuestas autores
Valoración global	Creo que el informe queda cojo si no se mete la perspectiva de los determinantes sociales de la salud, existe amplia evidencia, que tenemos que dársela a la máquina para que aprenda, de lo contrario corremos el riesgo de que el marco de pensamiento de la IA sea el dominante y, por lo tanto, quede sesgado hacia lo meramente clínico: alguna bibliografía a considerar: • Wilder ME, Kulie P, Jensen C, Levett P, Blanchard J, Dominguez LW, Portela M, • Srivastava A, Li Y, McCarthy ML. The Impact of Social Determinants of Health on • Medication Adherence: a Systematic Review and Meta-analysis. J Gen Intern Med.2021 May;36(5):1359-1370. doi: 10.1007/s11606-020-06447-0. Epub 2021 Jan 29. PMID: 33515188; PMCID: PMC8131473. • Karimi A, Burkhart RJ, Hecht CJ 2nd, Acuña AJ, Kamath AF. Is Social Deprivation Associated With Usage, Adverse Events, and Patient-reported Outcome Measures in Total Joint Arthroplasty? A Systematic Review. Clin Orthop Relat Res. 2023 Feb 1;481(2):239-250. doi: 10.1097/CORR.0000000000002394. Epub 2022 Sep 14. PMID: 36103392; PMCID: PMC9831197. creo que es necesario introducir también la perspectiva racial en el marco del informe no simplemente citando el el riesgo de sesgo, • Stabellini N, Dmukauskas M, Bittencourt MS, Cullen J, Barda AJ, Moore JX, Dent S, Abdel-Qadir H, Kawatkar AA, Pandey A, Shanahan J, Barnholtz-Sloan JS, Waite KA, Montero AJ, Guha A. Social Determinants of Health and Racial Disparities in Cardiac Events in Breast Cancer. J Natl Compr Canc Netw. 2023 Jul;21(7):705-714.e17. doi: 10.6004/jnccn.2023.7023. PMID: 37433439. • Karvonen KL, Anunwah E, Chambers Butcher BD, Kwarteng L, Mathis-Perry T, McLemore MR, Oh S, Pantell MS, Smith O, Rogers E. Structural	- Se ha añadido un párrafo en la discusión haciendo referencia a la necesidad de tener en cuenta los determinantes sociales de la salud en el diseño de sistemas de detección automatizada de RAM.

	Racism Operationalized via Adverse Social Events in a Single-Center Neonatal Intensive Care Unit. J Pediatr. 2023 Sep;260:113499. doi: 10.1016/j.jpeds.2023.113499. Epub 2023 May 19. PMID: 37211208.	
Resúmenes	Nada que aportar.	-
Introducción	Nada que aportar.	-
Metodología	Nada que aportar.	-
Resultados	Nada que aportar.	-
Discusión	Nada que aportar.	-
Conclusiones	Nada que aportar.	-
Anexos	Nada que aportar.	-
Otros comentarios	No representamos a ninguna asociación de pacientes, somos una asociación de profesionales. donde dice:” Asociación de Pacientes Polimedicados,” debe decir “Asociación Para las Prácticas Innovadoras en Polimedicación y Salud integrada en la Fundación para la Innovación, Investigación, Formación y el Desarrollo Comunitario (FÜNDEC). CIF. G04978730 N° Registro 2447 JUS www.fundec.org ”	- Se ha corregido el error

Revisiones de representantes de asociaciones de pacientes (RAP)

RAP_1

Sección del informe	Comentarios	Respuestas autores
Valoración global		-
Resúmenes		-
Introducción		-
Metodología		-
Resultados		-
Discusión		-
Conclusiones		-
Anexos		-
Otros comentarios	Por parte de la Federación Española de Diabetes (FEDE) consideramos adecuado el informe. Tan sólo nos gustaría conocer un aspecto, que consideramos no queda claro. Se habla como tema central, la identificación y, por ende, actuaciones ante eventos adversos producidos por medicamentos. No obstante, ¿cuándo se activa	- El informe se centra en la detección de RAM que ya han ocurrido, y no en la prevención de su ocurrencia, por lo que los sistemas evaluados se centran en la detección de

	ese control? ¿En el momento de la introducción de un nuevo medicamento o en el momento en el que se detecta un evento adverso producido por un medicamento nuevo?. Creemos que debería quedar más claro este protocolo de actuación.	síntomas para comprobar si están causados por medicamentos. En el caso de la prevención de RAM existen otros sistemas específicos como los sistemas de ayuda a la prescripción. Estos dos tipos de sistemas no son excluyentes, y pueden integrarse en un mismo servicio.
--	--	---

